

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 30, 2002, 09:34:06 ; Search time 1885.18 Seconds  
(without alignments)  
13009.846 Million cell updates/sec

Title: US-09-530-663B-5

Perfect score: 1172  
Sequence: 1 tftagattgtgtgaatg99.....agccgcgaatgacgcgaca 1172

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 1797656 seqs, 10463268293 residues

Total number of hits satisfying chosen parameters: 3595312

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : GenEmbl1.\*  
1: gb\_ba.\*  
2: gb\_htg.\*  
3: gb\_in.\*  
4: gb\_ov.\*  
5: gb\_ov.\*  
6: gb\_pat.\*  
7: gb\_ph.\*  
8: gb\_pl.\*  
9: gb\_pr.\*  
10: gb\_ro.\*  
11: gb\_scs.\*  
12: gb\_sy.\*  
13: gb\_un.\*  
14: gb\_vl.\*  
15: em\_ba.\*  
16: em\_fun.\*  
17: em\_hum.\*  
18: em\_in.\*  
19: em\_mu.\*  
20: em\_om.\*  
21: em\_or.\*  
22: em\_ov.\*  
23: em\_pat.\*  
24: em\_ph.\*  
25: em\_pl.\*  
26: em\_ro.\*  
27: em\_sts.\*  
28: em\_un.\*  
29: em\_vl.\*  
30: em\_htg\_hum.\*  
31: em\_htg\_inv.\*  
32: em\_htg\_other.\*  
33: em\_htgc\_inv.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No. Query Match length DB ID Description

1	1172	100.0	1172	8	AF041051	Populus t
2	826.8	70.5	4397	2	AY043494	Populus t
3	83	7.1	104992	8	AC005504	Plasmodiu
4	83	7.1	169546	2	AC004157	Plasmodiu
5	82.2	7.0	93368	9	AL133402	Human DNA
6	81.2	6.9	185699	2	AC021553	Homo sapi
7	81	6.9	236120	14	AF063866	Melanoplu
8	80.2	6.8	67970	3	PFMAL1P3	AL513330
9	80.2	6.8	202645	2	AC096550	Homo sapi
10	78.8	6.7	181486	2	AC096550	Homo sapi
11	78.4	6.7	14867	3	AE001398	Plasmodiu
12	78.4	6.7	189214	2	AC074035	Homo sapi
13	78.4	6.7	208319	2	AC108683	Homo sapi
14	78	6.7	86827	3	PFMAL1P5	AL034556
15	78	6.7	192731	9	AC026698	Homo sapi
16	77.8	6.6	172816	9	AC093899	Homo sapi
17	77.6	6.6	102401	2	AC094148	Rattus no
18	77.6	6.6	173480	9	AC090014	Homo sapi
19	77.4	6.6	245802	2	AC006279	Plasmodiu
20	77.2	6.6	104992	2	AC005504	Plasmodiu
21	77.2	6.6	169546	2	AC004157	Plasmodiu
22	76.8	6.6	158548	3	PFMAL1P2	AL034558
23	76.6	6.5	242159	2	AC091567	Homo sapi
24	76.4	6.5	176506	2	AC103500	Rattus no
25	76.2	6.5	53932	2	AC023371	Homo sapi
26	75.8	6.5	116696	3	PFMAL1P3	298547
27	75.8	6.5	186431	2	AC022281	Plasmodium
28	75.6	6.5	196355	9	AC099557	Homo sapi
29	75.4	6.4	17443	3	AC024806	Caenorhab
30	75.2	6.4	34548	6	AX349036	Sequence
31	75.2	6.4	178783	9	AC068139	Homo sapi
32	75.2	6.4	205130	2	AC105425	Homo sapi
33	75	6.4	44352	3	AF098501	Caenorhab
34	75	6.4	152359	9	AC093790	Homo sapi
35	75	6.4	182974	2	AC105148	Homo sapi
36	74.8	6.4	969	3	DMA132902	Drosophill
37	74.8	6.4	85779	8	SCE011856	Saccharom
38	74.8	6.4	167471	9	AC019211	Homo sapi
39	74.8	6.4	181792	9	AC098822	Homo sapi
40	74.8	6.4	253305	3	PFMAL1P7	AL034559
41	74.6	6.4	86827	3	PFMAL1P5	AL034556
42	74.6	6.4	88037	3	PFMAL1P8	Plasmodiu
43	74.4	6.3	35077	3	CEY38H8A	Caenorhab
44	74.4	6.3	118425	9	AL356005	Human DNA
45	74.4	6.3	189790	2	AC107420	Homo sapi

## ALIGNMENTS

RESULT 1  
AF041051 1172 bp DNA linear PLN 26-JUN-1998  
LOCUS Populus tremuloides clone PtAcLip 4-coumarate:COA ligase gene,  
DEFINITION Promoter region.  
ACCESSION AF041051  
VERSION AF041051.1 GI:3258638  
KEYWORDS  
SOURCE  
ORGANISM  
Quaking aspen.  
Populus tremuloides  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
Rosidae; eurosids I; Malpighiales; Salicaceae; Populus.  
REFERENCE  
Hu, W.-J., Kawacka, A., Tsai, C.-J., Lung, J., Osakabe, K., Ebinuma, H. and  
Chiang, V.-L.  
TITLE  
Compartmentalized expression of two structurally and functionally  
distinct 4-coumarate:COA ligase genes in aspen (Populus  
tremuloides)  
JOURNAL  
Proc. Natl. Acad. Sci. U.S.A. 95 (9), 5407-5412 (1998)  
MEDLINE  
98226828  
REFERENCE  
2 (bases 1 to 1172)  
AUTHORS  
Hu, W.-J. and Chiang, V.-L.

TITLE Direct Submission  
JOURNAL Submitted (06-JAN-1998) Plant Biotechnology Research Center, School  
of Forestry and Wood Products, Michigan Technological University,  
1400 Townsend Drive, Houghton, MI 49931, USA

FEATURES  
Location/Qualifiers

source  
1..1172  
/organism="Populus tremuloides"  
/db\_xref="taxon:3693"  
/clone="P14CL1P"  
/tissue\_type="leaf"  
1..>1172  
/gene="4-coumarate:COA ligase"  
promoter  
1..1172  
/gene="4-coumarate:COA ligase"  
BASE COUNT 399 a 224 c 181 g 368 t  
ORIGIN

Query Match 100.0%; Score 1172; DB 8; Length 1172;  
Best Local Similarity 100.0%; Pred. No. 4.6e-182;  
Matches 1172; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ttgtagattggtggaatggatcattccctaatgacggtggaacacaa 60  
DB 1 tctagcattggtggaatggatcattccctaatgacggtggaacacaa 60  
QY 61 gcaaaagagaagttagtcctcctctctatatatatatatatgcatgagagacc 120  
DB 61 gcaaaagagaagttagtcctcctcctctctatatatatatatatgcatgagagacc 120  
QY 121 atggtcatgatagaagttatagaggtagtctgattgagatagtcacgacactagttt 180  
DB 121 atggtcatgatagaagttatagaggtagtctgattgagatagtcacgacactagttt 180  
QY 181 tttgttggtggttct 240  
DB 181 tttgttggtggttct 240  
QY 241 tttgttggtggttct 300  
DB 241 tttgttggtggttct 300  
QY 301 agtgcataatgacacactgacactgacactgacactgacactgacactgacactgacact 360  
DB 301 agtgcataatgacacactgacactgacactgacactgacactgacactgacactgacact 360  
QY 361 aaaaactaaataaaagattagattatataatattagttatcaaggttggtgct 420  
DB 361 aaaaactaaataaaagattagattatataatattagttatcaaggttggtgct 420  
QY 421 aatcaattatataataaagcagtagtatttgataatataatataatataatatttg 480  
DB 421 aatcaattatataataaagcagtagtatttgataatataatataatataatatttg 480  
QY 481 attgaatgaactcaattacatcacaaaaaacttaataatataatctatgagat 540  
DB 481 attgaatgaactcaattacatcacaaaaaacttaataatataatctatgagat 540  
QY 541 aattgagaatataatgattacacttaactgagatttctctataaaaaacagcga 600  
DB 541 aattgagaatataatgattacacttaactgagatttctctataaaaaacagcga 600  
QY 601 taattgggtagaattcaagcattatcaaacctgcccagagcaatttaaatatt 660  
DB 601 taattgggtagaattcaagcattatcaaacctgcccagagcaatttaaatatt 660  
QY 661 aattatatttttcttaataaagcacttcttaattgtaaaatataatgcttaaacactaa 720  
DB 661 aattatatttttcttaataaagcacttcttaattgtaaaatataatgcttaaacactaa 720  
QY 721 taataaattatttctgatactcttgaggcagtagtgagaggtgcgacaaataatagt 780  
DB 721 taataaattatttctgatactcttgaggcagtagtgagaggtgcgacaaataatagt 780

QY 781 gcaataataatagtgattggtctgtgaaaagacagtgagagcaaacactctc 840  
DB 781 gcaataataatagtgattggtctgtgaaaagacagtgagagcaaacactctc 840  
QY 841 tcaagtcataaaggcatttcaacaacccaatggaacccaccgctcccgca 900  
DB 841 tcaagtcataaaggcatttcaacaacccaatggaacccaccgctcccgca 900  
QY 901 ttaaaatcttaattctcaccacccaactcccaagttcttcaacaaagcaatgatt 960  
DB 901 ttaaaatcttaattctcaccacccaactcccaagttcttcaacaaagcaatgatt 960  
QY 961 ttaaaatcttaattctcaccacccaactcccaagttcttcaacaaagcaatgatt 1020  
DB 961 ttaaaatcttaattctcaccacccaactcccaagttcttcaacaaagcaatgatt 1020  
QY 1021 ttaaaatcttaattctcaccacccaactcccaagttcttcaacaaagcaatgatt 1080  
DB 1021 ttaaaatcttaattctcaccacccaactcccaagttcttcaacaaagcaatgatt 1080  
QY 1081 tttgttctgaagttct 1140  
DB 1081 tttgttctgaagttct 1140  
QY 1141 ctctgcatctttagcccgcaatgagcgaca 1172  
DB 1141 ctctgcatctttagcccgcaatgagcgaca 1172  
RESULT 2  
AY043494 4397 bp DNA linear PLN 17-SEP-2001  
LOCUS Populus tomentosa 4-coumarate:COA ligase gene, complete cds.  
DEFINITION AY043494  
ACCESSION AY043494  
VERSION AY043494.1 GI:15636676  
KEYWORDS  
SOURCE  
ORGANISM  
Populus tomentosa.  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
Rosidae; eurosids I; Malpighiales; Salicaceae; Populus.  
REFERENCE  
1 (bases 1 to 4397)  
Lu, H., Zeng, Q., and Jiang, X.  
Genomic DNA sequence of 4-coumarate:COA ligase gene from Populus  
tomentosa  
JOURNAL Unpublished  
REFERENCE 2 (bases 1 to 4397)  
AUTHORS Lu, H., Zeng, Q., and Jiang, X.  
TITLE Direct Submission  
JOURNAL Submitted (01-JUL-2001) Biology, Beijing Forestry University,  
Beijing Qinghua East Road No. 35, Beijing 100081, China  
FEATURES  
Location/Qualifiers  
1..4397  
/organism="Populus tomentosa"  
/db\_xref="taxon:118781"  
1..1136  
/product="4-coumarate:COA ligase"  
/join(1137..2136,3321..3519,3644..3857,3957..4059,  
4174..4284)  
/codon\_start=1  
/product="4-coumarate:COA ligase"  
/protein\_id="AA02144.1"  
/db\_xref="GI:15636677"  
/translation="MNPEEFIFRSKLPDIYIPKMLPLHSYLENLNHSKPKCLING  
ANGDIYAYADVELTARRVAGSLIGIOGDIIMFLSSPEFVAFGLASRGAIT  
AANPSTPAELAKHAKASRAKLITQACYEKVPFARSDVWKCVAQSPGCLHFS  
ELTQADENAPDOVIDSPDDVVALPYSSGFTGIPKGMVLTIRKGLITSVQVQDGNPL  
VFHSRDVILCVLPMFHITIALNSIMLCIRVGCASLIMPFTEGLSLGLIEKKVSIAP  
VVPFVMSIASFSPDLDKHDLSLRMKISGAGLUGLEDTYNAKFPQARLGGGTGTE  
AGPVLAMCLAFKPEPFDIPGACGTVVNAEMKIVDPETGASLPNPGEICINDOI



[illegible]

LOCUS	AL133402/c	93368 bp	DNA	linear	PRI 24-JUL-2000
DEFINITION	Human DNA sequence from clone RPS-1077H22 on chromosome 6. Contains				
FEATURES	<p>source</p> <p>1. 169546 /organism="Plasmodium falciparum" /db_xref="taxon:5833" /chromosome="12" /clone="PEYAC293" /clone="3D7"</p>				
BASE COUNT	69871 a 15381 c 15705 g 68389 t 200 others				
ORIGIN					
Query Match	7.1%: Score 83; DB 2; Length 169546;				
Best Local Similarity	45.3%: Pred. No. 6.1e-05;				
Matches 302; Conservative	0; Mismatches 365; Indels 0; Gaps 0;				
QY	133	aaggtataagaggtacgttcgtatcgatagatgtccacgactagtttcgtgtgtga	192		
Db	86291	AATATATATATTTATTTATATATATATATATTTAAATTAATTAATTTATTTATTAATTA	86232		
QY	193	ttccatcagatgacgcgaaaaatttatcatatataataaagaataatgactgattacc	252		
Db	86231	TTTAAATTAATAATTAATAATTAATTAATTTATTTATTTATTAATTAATTAATTAATAA	86172		
QY	253	cctcgtaatttggtaaaatagataaacaacgcctcaatgtgaggtgacacctgtcaaat	312		
Db	86171	TTTATTTATTTATTTATTAATTAATTTTAAATTAATAATTAATTAATTTATTTATTAATTA	86112		
QY	313	accactcgactggggcactgtgattcttccaatcacaactccaatttgaacaaatc	372		
Db	86111	AAATATTTAAATTAATAATTTAAATTTATTTATTTATTTAAATTAATAATTAATTAATTAAT	86052		
QY	373	aaaaaagctttagatcatataaattttagttaatcaccgggttggctacatattata	432		
Db	86051	AATATATATTTATTAATTTATTTAAATTAATAATTAATTTCAATTTATTAATTTATTAATTT	85992		
QY	433	ttaattaaacagatagtaattttagtaataatttaataaattttagtattgattgaatgaac	492		
Db	85991	TTAAATTAATTAATTAATAATTAATAATTAATTTAAATTTAAATTTAAATTAATAATTAATTA	85932		
QY	493	tcaattacatccaaaaaacctcaatcaaatatatactctatgtgataatatttgaata	552		
Db	85931	TTAAATTTAAATAATTTAAATTAATAATTAATTAATTTATTAATAATTTAAATTAATTAATTA	85872		
QY	553	taaatgattaacctttaaactcgcggttcctctataaaaaacgcgataaatttggcctag	612		
Db	85871	TTTTTAAATTAATAATTTCTTAATTTATTTTATTTACATTTTATTAATAATTAATTTTATTT	85812		
QY	613	alttaacagctatcttcaaactggccaggaacattataaataataatcatattatttc	672		
Db	85811	AAAAAAATTTATTTAAATTAATAATTTAAATTTAAATTTAAATTTAAATTAATAATA	85752		
QY	673	ttcttaataaagcactccctaatttgttaaatatctgcttcaaacctcaataataaattta	732		
Db	85751	TTTAAACAAATTAATTAATAATTAATTAATTAATTTAAATTTAAATTTAAATTTAAATTTAT	85692		
QY	733	tttgtctactcttggccagtagtgaggtgcgtgcgcaataaattatgacataaata	792		
Db	85691	TTATATATTTATTAATTAATTTATTAATTAATTTAAATTTAAATTTAAAGAAAGATTAATTT	85632		
QY	793	atggatt 799			
Db	85631	ATACCTT 85625			



ACCESSION	VERSION	KEYWORDS	SOURCE	ORGANISM	REFERENCE	AUTHORS	TITLE	JOURNAL																									
AL133402	AL133402.10	GI:8039186	HTG.	human.	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.	Submitted (18-JUL-2000) Sanger Centre, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries: humquerry@sanger.ac.uk	requests: clonerequests@sanger.ac.uk	On May 23, 2000 this sequence version replaced gi:7939106. During sequence assembly data is compared from overlapping clones. Where differences are found these are annotated as variations together with a note of the overlapping clone name. Note that the variation annotation may not be found in the sequence submission corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above. This sequence has been finished according to sequence map criteria as follows. An attempt is made to resolve all sequencing problems, such as compressions and repeats, but not necessarily within known annotated human repeat sequence elements (e.g. Alu). Where the sequence is ambiguous, there is an annotation using the 'unsure' feature key.																									
FEATURES	source	Location/Ouallifiers	1..93368	/organism="Homo sapiens"	/db_xref="taxon:9606"	/chromosome="6"	/clone="RP5-1077H22"	/clone_id="RPC1-5"	231..537	/note="AluY repeat: matches 1..301 of consensus"	638..892	/note="AluX repeat: matches 39..303 of consensus"	1142..1329	/note="MIR repeat: matches 30..235 of consensus"	1386..1680	/note="AluSg repeat: matches 1..304 of consensus"	complement(3971..4298)	/note="match: GSS: Em:AQ720689"	4299..4722	/note="L1PBa repeat: matches -1546..-1136 of consensus"	4721..6631	/note="L1PBa repeat: matches -134..-2762 of consensus"	6632..6939	/note="AluY repeat: matches 3..310 of consensus"	6940..7827	/note="L1PBa repeat: matches 2762..3638 of consensus"	7856..9561	/note="L1PB2 repeat: matches 4121..5791 of consensus"	9565..9598	/note="match: GSS: Em:AQ984702"	9599..9750	/note="4 copies 38 mer 71% conserved"	9604..9739

misc_feature	/note="68 copies 2 mer ta 73% conserved"
misc_feature	9708..9831
misc_feature	/note="match: SFS: Em:G34561"
misc_feature	complement(9730..9849)
misc_feature	/note="match: SFS: Em:252156"
misc_feature	9764..9824
misc_feature	/note="match: GSS: Em:AQ014122 Em:AQ598220"
misc_feature	complement(9764..9823)
misc_feature	/note="match: GSS: Em:AG019010 Em:AG026881"
misc_feature	complement(9764..9824)
misc_feature	/note="match: GSS: Em:AQ933387 Em:AQ023541"
misc_feature	9772..9824
misc_feature	/note="match: GSS: Em:A2110479"
misc_feature	complement(9796..9835)
misc_feature	/note="match: GSS: Em:A2002912"
misc_feature	9796..9835
misc_feature	/note="match: GSS: Em:AL214116 Em:AQ341248"
misc_feature	9796..9836
misc_feature	/note="match: GSS: Em:AQ548435"
misc_feature	complement(9796..9828)
misc_feature	/note="match: GSS: Em:AQ324044 Em:AQ373503"
misc_feature	complement(9796..9824)
misc_feature	/note="match: GSS: Em:AQ598220"
misc_feature	9796..9823
misc_feature	/note="match: GSS: Em:AQ745266"
misc_feature	9796..9824
misc_feature	/note="match: GSS: Em:AQ014122"
misc_feature	9797..9828
misc_feature	/note="match: GSS: Em:AQ745589"
misc_feature	9797..9824
misc_feature	/note="match: GSS: Em:AQ014122"
misc_feature	9797..9841
misc_feature	/note="match: GSS: Em:AQ745266"
misc_feature	9803..9824
misc_feature	/note="match: GSS: Em:AQ598220"
misc_feature	9806..9841
misc_feature	/note="match: GSS: Em:AG026544"
misc_feature	9807..9841
misc_feature	/note="match: GSS: Em:A2028723"
misc_feature	9842..10202
misc_feature	/note="L1P82 repeat: matches 5782..6154 of consensus
misc_feature	10724..11003
misc_feature	/note="AluJo repeat: matches 1..289 of consensus"
misc_feature	11341..11501
misc_feature	/note="MERB repeat: matches 5..173 of consensus"
misc_feature	12016..12384
misc_feature	/note="WR1B repeat: matches 1..387 of consensus"
misc_feature	13221..13307
misc_feature	/note="HA1I repeat: matches 550..638 of consensus"
misc_feature	13294..13693
misc_feature	/note="HA1I repeat: matches 595..1036 of consensus"
misc_feature	14165..14463
misc_feature	/note="AluI repeat: matches 1..296 of consensus"
misc_feature	14897..15051
misc_feature	/note="L1MC repeat: matches 267..420 of consensus"
misc_feature	15208..15761
misc_feature	/note="MER6B repeat: matches 1..486 of consensus"
misc_feature	17062..17326
misc_feature	/note="AluIb repeat: matches 28..294 of consensus"
misc_feature	complement(19549..20029)
misc_feature	/note="match: GSS: Em:AQ734906"
misc_feature	join(<21618..21754,25301..>25566)
misc_feature	/gene="dJ1077H2.1"
misc_feature	/note="match: ESTs: Em:AA960797"
misc_feature	/evidence="not_experimental"
misc_feature	/product="dJ1077H2.1 (putative novel protein)"
misc_feature	join(21618..21754,25301..25566)
misc_feature	/gene="dJ1077H2.1"
misc_feature	24053..24358
misc_feature	/note="MER3 repeat: matches 7..323 of consensus"
misc_feature	25436..25647
misc_feature	/note="MIR repeat: matches 43..246 of consensus"
misc_feature	26551..26628

	repeat_region	/note="MIR repeat: matches 44. .121 of consensus" 27163. .27450 /note="AluSc repeat: matches 1. .288 of consensus" 28099. .28569 /note="MLTID repeat: matches 2. .505 of consensus" 32415. .32450 /note="L2 repeat: matches 2677. .2710 of consensus" 34405. .34634 /note="LMB8 repeat: matches 5857. .6120 of consensus" 34944. .36430 /note="L2 repeat: matches 62. .1680 of consensus" 36549. .36873 /note="L2 repeat: matches 2411. .2750 of consensus" 37369. .37635 /note="MER33 repeat: matches 22. .294 of consensus"
	repeat_region	37654. .37931 /note="LIMC5 repeat: matches 7507. .7777 of consensus"
	repeat_region	38000. .38111 /note="28 copies 4 mer tata 67% conserved" 38001. .38112 /note="56 copies 2 mer at 67% conserved"
	repeat_region	39562. .39844 /note="Alut repeat: matches 1. .281 of consensus"
	repeat_region	41294. .41412 /note="MIR repeat: matches 53. .171 of consensus"
	repeat_region	41413. .41819 /note="MSR repeat: matches 1. .426 of consensus"
	repeat_region	41820. .41874 /note="MIR repeat: matches 171. .225 of consensus"
	repeat_region	42349. .42490 /note="MIR repeat: matches 34. .185 of consensus"
	misc_feature	43137. .43480 /note="match: GSS: Em:AQ430150"
	repeat_region	44406. .44524 /note="MIR repeat: matches 21. .146 of consensus"
	repeat_region	45108. .45474 /note="MER74A repeat: matches 132. .528 of consensus"
	repeat_region	45548. .45644 /note="MIR repeat: matches 56. .147 of consensus"
	repeat_region	46184. .46239 /note="14 copies 4 mer agaa 73% conserved"
	repeat_region	46186. .46375 /note="5 copies 38 mer 65% conserved"
	repeat_region	46228. .46383 /note="4 copies 39 mer 76% conserved"
	repeat_region	46452. .46565 /note="3 copies 38 mer 75% conserved"
	repeat_region	46474. .46565 /note="23 copies 4 mer gagag 65% conserved"
	repeat_region	46483. .46558 /note="38 copies 2 mer gg 67% conserved"
	repeat_region	47326. .47676
	Query Match	7.0%; Score 82.2; DB 9; Length 93368; Best Local Similarity 47.7%; Pred.No.9.5e-05;
	Matches 277; Conservative	0; Mismatches 298; Indels 6; Gaps 1.
Dy	210 aaatttataatatataatgaataatgatgattatcttcgtgaatttgtaa	269
Dd	77802 ATACATTATATAATTATTAATTTAAATATAATGTAATATAATTAATATA	77743
OY	270 atagaattaacaogcactgcgaggcgccggttgccaatgcacccgcgaactgggac	329
Dd	77742 TTATTTATAAAATATATATCTGTATATATAATTAATTAATTAATATAATATA	77683
OY	330 atggtagattttaaacaccacccaatttgaaaaactaaaattaaaagaattagalta	389
Dd	77682 TTAGATATAAATATATTTATTAATTTATATATAATAATATAATTGGATATATAATTAAT	77623
OY	390 ttaattattgggtcaattcacgggttgtcctaatacatcataatcaataaagatalg	449
Dd	77622 ATATATATAAATAATATAATT-----GTATATATATTAATTAATATAATATAAATAATATA	77569

OY	450	atttttgtaatttbaatcaaaatttatctgggttcggaagcacaataacccaanaa	509
Db	77568	ATTGATATTAATATAATTTATATATATATAAAATATTAATGTGATATTAATATAA	77509
OY	510	aacctbaalcaaatataactctatgltgatataactlagaalacataaagtataacctta	569
Db	77508	TATATTAATATATAAATAATTAATGAATATATTAATTAATTAATTAATAATTAATA	77449
OY	570	aatcccgagttccctctaataaaaacagataattgggcctgattcaacgctattt	629
Db	77448	TAAATGTAATATAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT	77389
OY	630	caaactggccsgacaaltataataaattaatactatctttccctaagaaccctc	689
Db	77388	ATATATTAATATAAAAAATATAATTAATGAATATAAAAATAATTAATTAATATAAAAAA	77329
OY	690	ctaatgtctaataatatgctcaaacacatacaataaattatctgcatcttggca	749
Db	77328	TATATATGTAATATAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT	77269
OY	750	gtaggtagaggtgctgcacaataaattagycataaata	790
Db	77268	ATTATATATTAATTAATAAATATAAATTAATGAATATAAATA	77228

RESULT	6
AC021553	
LOCUS	AC021553 185699 bp DNA linear HTG 05-JAN-2001
DEFINITION	Homo sapiens chromosome 15 clone RP11-709B3 map 15, *** SEQUENCING
ACCESSION	AC021553
VERSION	AC021553.13 GI:15706183
KEYWORDS	HTG; HTGS_PHASE2; HTGS_FULLTOP; HTGS_ACTIVEFIN.
SOURCE	human
ORGANISM	Homo sapiens
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS	Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
TITLE	1 (bases 1 to 185699)
JOURNAL	Birren,B., Linton,L., Nusbbaum,C. and Lande,E.
REFERENCE	Homo sapiens chromosome 15, clone RP11-709B3
AUTHORS	Unpublished 2 (bases 1 to 185699) Birren,B., Linton,L., Nusbbaum,C., Lande,E., Abraham,H., Allen,N., Anderson,S., Baldwin,J., Barna,N., Beckerly,R., Beda,F., Boguslavsky,I., Boukhalter,B., Brown,A., Burkett,G., Castle,A., Chapel,Y., Colangelo,M., Collins,S., Collamore,A., Cooke,P., DeRellano,K., Dewar,K., Domino,M., Doyle,M., Fenestor,J., Ferreira,P., FitzHugh,W., Forrest,C., Gage,D., Galagan,J., Gardina,S., Grant,G., Hagos,B., Heathford,A., Horton,L., Howland,J.C., Johnson,R., Jones,C., Kamm,L., Karatas,A., Klein,J., Landers,T., Lehoczy,J., Levine,R., Lien,C., Liu,G., Locke,K., McDonald,P., Marguis,N., McEwan,P., McGurk,A., McKernan,K., Mcneelers,R., Meldrum,J., Meneus,L., Morrow,J., Naylor,J., Notman,C.H., O'Connor,T., O'Donnell,P., Olivar,T.M., Peterson,K., Pierre,N., Pisanal,C., Pollara,V., Raymond,C., Riley,R., Rottman,D., Roy,A., Santos,R., Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J., Testaye,S., Theodore,J., Tirrell,A., Vassiliy,H., Viel,R., Vo,A., Wu,X., Wyman,D., Ye,W.J., Zimmer,A. and Zody,M.
TITLE	Direct Submission
JOURNAL	Submitted (16-Jan-2000) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA On Sep 20, 2001 this sequence version replaced gi:15148276. All repeats were identified using RepeatMasker: Smit, A.F.A. & Green, P. (1996-1997) <a href="http://ftp.genome.washington.edu/RM/repeatmasker.html">http://ftp.genome.washington.edu/RM/repeatmasker.html</a>
COMMENT	----- Genome Center Center: Whitehead Institute/ MIT Center for Genome Research Web site: <a href="http://www.seq.wi.mit.edu">http://www.seq.wi.mit.edu</a> Contact: <a href="mailto:sequence_submissions@genome.wi.mit.edu">sequence_submissions@genome.wi.mit.edu</a> ----- Project Information Center project name: I5662



gene  
LTKLYVAVIIONLEVI"  
complement(7431. .8171)  
/gene="MSV005"  
complement(7431. .8171)  
/gene="MSV005"  
/codon\_start=1  
/product="ORF MSV005 hypothetical protein"  
/protein\_id="AAC97860.1"  
/db\_xref="GI:4049900"  
/translation="MKMKILFKDYNTFKIIEYIKSLDGCYNLKKKLNIDDKLSIKIK  
LTCYNPISSECVLPISIKFIKFIYTLSEIILILNLNYIKQMLNIPRAKHNAYKKG  
KCELTIFQNOPEICELCYRKKNVMSCHDIDFVELDKCNKCNKFNKISYAYAG  
VLAKKVIKIKTIGNDYISNICECKYIYLFKRLPYTRCIGNNFIIDSELNTKFC  
HYCYFYCYWNGYIWIYIPLFDNYNG"  
8831. .9058  
/gene="MSV006"  
8831. .9058  
/gene="MSV006"  
/codon\_start=1  
/product="ORF MSV006 hypothetical protein"  
/protein\_id="AAC97608.1"  
/db\_xref="GI:4049648"  
/translation="MFEISFSDMNIILINDSKNVKFLIQEIHFNISINDISVHASY  
IPEELIPTNSNILLINSKLNKSIISGLQ"  
complement(9825. .11327)  
/gene="MSV008"  
complement(9825. .11327)  
/gene="MSV008"  
/codon\_start=1  
/product="ORF MSV008 leucine rich repeat gene family  
protein, similar to Amsacta moorei entomopoxvirus Q3 ORF  
SM:P28854"  
/protein\_id="AAC97859.1"  
/db\_xref="GI:4049899"  
/translation="MNNINLYCCKNIILKYLDNDOKNNGIYELFKYYNCILKPIYIY  
KIFYNYNLEYIKNMCSDCNTDFKELESLINLELIDSYKNSNIYCKRSLIKMLY  
NCKEKKIDFKLESLINDELIDSYKNSNIYCKRSLIKNCKNECKITDFKPE  
SLINEILIDSYKNSNIYCKRSLIKNCKNECKITDFKPE  
YSNIDCKLPVSLKLNCEFYCYDPFKLEPLNMNKLINISYKNSNIDCKLPVSL  
IKLNGCYCYDFKLEPLINLQKLNISNSYSNCKSLISLKFESGCDITNFK  
LEFLRINLETLITVMSNSISICKPLSLINLNFIEDYDINIKFELINLKLSTATS  
SNSSNRNISCIPLISLYVDLYSCHIKNFKLEPLINLKNKINISDNGDSNRSKRLPI  
SLIEIYCSCKNIKNFLOPLINLOKLDISVDFSENLNHNFPETN"  
10282. .10476  
/gene="MSV007"  
10282. .10476  
/gene="MSV007"  
/codon\_start=1  
/product="ORF MSV007 hypothetical protein"  
/protein\_id="AAC97609.1"  
/db\_xref="GI:4049649"  
/translation="MYIVNSKILNRSKLNKLFVISOYSDMLINDIGNLHDIPEYELE  
ELIHPCKIKNSKLNKLSYI"  
complement(11374. .12138)  
/gene="MSV009"  
complement(11374. .12138)  
/gene="MSV009"  
/codon\_start=1  
/product="ORF MSV009 leucine rich repeat gene family  
protein, similar to Amsacta moorei entomopoxvirus Q3 ORF  
SM:P28854"  
/protein\_id="AAC97858.1"  
/db\_xref="GI:4049898"  
/translation="MNNLTYEYKELKLNLDNLTKHLAARNKLYRKYNSIINPLHBE  
NHIKYANPKHVKKLOCSGIPNKLINLELIDSYKNSNIYCKRSLIKMLY  
ITDFKLENLVLEKFIINKNKSNISDCKLPISITIECCNCKETIDPKFLEDLNKLK  
KLIYSGNKNKSNCHCNLPKILTELCYSTCNIIIDFELNVLNKLKLTFTGNGSNISIS  
CELPRKLIFVCYNINIDPFLFKSSNNLRITISN"  
complement(12364. .14199)  
/gene="MSV010"  
complement(12364. .14199)  
/gene="MSV010"  
/codon\_start=1

[illegible]

Db 228214	AATTCATATATTATTAATTAAATTAACGATTTTGTGTAATATTAATATCATTTTCA	228159
OY	413 ggtggcctaacaattacttaattaacaacgatgialttttgtaataattaataaat	472
Db 228158	TTTGAAAATATCGTAATATTAATTTATTACATATATCTAATTCGTTTAAATTAATTAATTT	228099
OY	473 ttattggatttgaaatgaactcaactaatacaaaaaaaccaatcaaatlaactta	532
Db 228098	TTTAAAGATTTAGCATTTATTATTTATTTGCAATTTACATTAATTTAAATTTATTAATTAATGATTTT	228039
OY	533 tctgataataattgagaataataaalgatbaaccttlaaatctcgagtttctctlaaaa	592
Db 228038	GCTAATATTATTAATATCATTTGCAATTTGGAATAATCTAATATCCAAATTTATAGATTAATTT	227979
OY	593 aacacgtataattgggcctagatattaacsctatattccaacctggccagacaattatta	652
Db 227978	AATTCCTGTAATTAATTAATAATTTTGGAGATCTCATCATTTTCAATTAATTAACATTTACTCA	227919
OY	653 aaattaataataattatcttctcaataaagcaacttccbaattcgtlaaataataatgctta	712
Db 227918	ATATTTAATTTAATTAATTAATGATCTCGTAATTCGTAGCACCATATGTAATTTATTATTAGAT	227859
OY	713 aacctataataaataattatcttgctatccttggcagtgagtgagggctgcgaacaat	772
Db 227858	AATATCTAATTAATTTGTAAATTAATTAATAATTTCTTAAAAAATTTAAAAGATTTAATTAATTAAT	227799
OY	773 aaattagctcaaaataataaalgatggttggctgc 807	
Db 227798	GAACGCAATTTAAATTAATTAATTAATTTGGTAATTTT 227764	

FEATURES	source
LOCUS	67970 bp DNA linear INV 15-DEC-1999
DEFINITION	Plasmodium falciparum MAL1P3, complete sequence.
ACCESSION	AF031746
VERSION	AF031746.9 GI:6594243
KEYWORDS	HTG.
SOURCE	malaria parasite P. falciparum.
ORGANISM	Plasmodium falciparum
REFERENCE	Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
AUTHORS	1 (bases 1 to 67970) Bowman, S., Churcher, C., Harris, B., Harris, D., Lawson, D., Quail, M. and Barrell, B.
TITLE	Direct Submission
JOURNAL	Submitted (24-SEP-1998) P. falciparum Genome Sequencing Consortium. The Sanger Centre, Wellcome Trust Genome Campus, Hinxton, Cambridge CB10 1SA, UK
COMMENT	On Dec 16, 1999 this sequence version replaced gi:5763807. For more information about this sequence or the Malaria Project, see <a href="http://www.sanger.ac.uk/Projects/P_falciparum">http://www.sanger.ac.uk/Projects/P_falciparum</a> . IMPORTANT: This sequence is unfinished and does not necessarily represent the correct sequence. Work on the sequence is in progress and the release of this data is based on the understanding that the sequence may change as work continues. The sequence may be contaminated with foreign sequence from E.coli, yeast, vector, phage etc.
FEATURES	Location/Qualifiers
source	1..67970
	/organism="Plasmodium falciparum"
	/strain="3D7"
	/db_xref="taxon:5833"
	/chromosome="1"
	complement(join(1748..2598,2748..2848,2990..3276))
	/gene="MAL1P3.01"
	complement(join(1748..2598,2748..2848,2990..3276))
	/gene="MAL1P3.01"
	/note="MAL1P3.01, conserved hypothetical protein, len: 412 aa, similarity: UPI0006 family eg to YB1055C/YB10512/YB10511, YPF5_YEAST (418 aa), fasta scores: opt: 316, E(): 1.1e-12, (33.28 identity in 271 aa overlap)"
	/codon_start=1

misc\_feature  
/product="conserved hypothetical protein, UF0006 family"  
/protein\_id="CAB63556.1"  
/db\_xref="GI:6594244"  
/translation="MKLVEHYIKYINLVETIISILFKLSNLSLKLYNDLRISTVAKYKV  
IOLIKRSNLKKNHNRKMEKDESPIDIGSLM/TGKMPGVYNSKHENDLONVLRNK  
NNVVKITITICGLAEIDKSLKICTGYDEKFLVLSAGYVNCYCFIDIKHEEKE  
IIAKKEYEPIITFKNEQVNSKMBNGKTKCDEKGMNNINELLEKNIDTIPGYE  
NKKEYLEENLNKNIITKYPNIVCGIGELGDFDLVFCSTKIQTIKYTFOLKVOVN  
LPMFLHMRKSTPEFKVIDIYKFLFENKGVYHSFTQEKEDIVHIVONYNKLVIANGV  
SLKLEENINNAVKIPLMLILLETDPAWCGVKTHASKEYIKDTYERKRAYTNLKKIN  
IKGCDNITFKERNPYVIA"  
complement(2599..2610)  
/gene="MALP3.01"  
/note="potential splice acceptor sequence"  
complement(2744..2747)  
/gene="MALP3.01"  
/note="potential splice donor sequence, atg/gttaa"  
complement(2844..2861)  
/gene="MALP3.01"  
/note="potential splice acceptor sequence"  
complement(2984..2989)  
/gene="MALP3.01"  
/note="potential splice donor sequence, aaa/gtaaa"  
5005..5496  
/gene="MALP3.02"  
5005..5496  
/gene="MALP3.02"  
/note="MALP3.02, hypothetical protein, len: 163 aa,  
contains possible signal sequence"  
/codon\_start=1  
/product="hypothetical protein, MALP3.02"  
/protein\_id="CAB63557.1"  
/db\_xref="GI:6594245"  
/translation="MKLNNRFVVCPIILIEFFLNSVVLGNNRRNNINPHETNNAK  
ARKRLINSGINSIKLDNGDELKIKNDKHHDSQKWKQSFISINLEEKQSQDRLR  
KQOEINBANTKIIEIROEFTIINNDIEINATRYLENPFELTIQSFKSLIDIOS  
LNH"  
8020..10389  
/note="possible cen1, region of very high [A+T] content"  
14884..20352  
/gene="MALP3.03"  
14884..20352  
/gene="MALP3.03"  
/note="MALP3.03, putative ABC transporter, len: 1822 aa"  
/codon\_start=1  
/product="putative ABC transporter"  
/protein\_id="CAB63558.1"  
/db\_xref="GI:6594246"  
/translation="MTTYKENVGINSKGNKKKSCONISFLNPLSPDMIRPLINDLIK  
GIOELPNICRFVDYASKLEENLRDIEVEDSEFYSEKSSNEHVLHHCNSDASEE  
KVVYVYVHNIMLSLTKTEKFPRIILISFTYILEPTLVIGLGFIDYMRILEGOKIPV  
YISFLKDRFVSGLVYVIMFPHLFFELLALFYTHLPTINLAKYSLMFLYKINCSNN  
NHLQNDPAINYTRKFFSSOTEIDELSRDFLSIGKNASSSGIKNNKNINDNNKVFVN  
DYIINFISKTRMEKDSLNEKNSLPPNVIYVIMDSVPSEVFEVTCINLFENFVKLFE  
MSFYVFKHIGSGNSVGIAMLSIALYSAIMLFEPLSPFLSKYLIYDQRIDNNHVL  
KEFKYILFENKSSFAFKYINIEPKEMKYCKIRLYLSNIGVYISISDIIEVYIFET  
LYKDLKNKEELKEFTSIIMPLYVYKILISVANNPPLVNNMEGIVINILRYNNINOH  
LYYNDIKNFKETRYNEDNYIVYDKTFLQNDENTSDDQSNHLNKLKAYIKRKLNN  
MKRYTFEKNKTYINIKQILSLGLNVDNTRNKICPOEHSNSTYVNNSSHINE  
KKEEYENIHSSGMSSEKFAKKNNNYIILKLENSGSLGYDKKCONDHLKNIINN  
LKRNSLAIIGNVGSGSAFEPKSLIGDNPMMHNGIYIENFKRPIILYVQDQSNL  
NIRNSLIGENEYNIYKTYTILQSLANDLSITIEBGMDYIINDNHSKSGKVCICLA  
NRLAYEYIHHMKLCTDYERKLIQPELIDKPLINKNKISVYNNKSKLVYNIIPFEN  
YQGLKLEMDNVOYIILDDIFPSLDRPSIKSTFSLKEDNISFPKQNSITISMNS  
TLDNLELIDNVOYEVNIEFDQTKLTKGNISETKNNNLITESHGNSLWLT  
IDYTRKLFDEVELNHVKSNMKIKRAYEVKAGTESVSEFIDISNEYIIRKKMKYK  
KKEHNNKNNKNNNNNNNSKDDHINIMANDNHRNYDINLGNPSTDSPPVSSLGNE  
YLDYDTSSNSKDEIVKPLKXDTHEENKSSMPKVSNNMINSNFEYEDNSSSS  
FEGSISLEFYLMYFOQGVFVLITSVIEMLISITFDLTKRPFVLMMSIISKNKHESD  
TTLQOVORILEVFLPIISLVTSIGCSMTIYNTISAIRVHNHILYSLINAPLYE  
YANNNGININRTIDISAFDYGFLKRIYKAFIFPRCLSSLLITVIMRODIFPPV  
ILILYEFKFRSGCKEAOURLYLSCHPLTNYISMLSGKNIINITYKNTYHLDVE  
HININRISYEFKMLINIMASLYIKIFILLPLTIVYIMHPHLYASGIIKLYEKYVVI

LSLTGVCISFSARLGIILIKFLCDYTHIEKMCVORLEEFPAKISNKENMSKNENEL  
NVITTOYKRENNENISDKISAIYEYKNNVSLISSODDESKKKYGRKFNVYVYK  
KPILVNGTYKRIIDEPSIKNINMYALKNOKIGIVGSGAKSSTILISGILINISOG  
KITVEGRDIRTYNNKKGEDSTIGITIAOSSFPYNNINRIFIDPYNNPDDDELVAHLKN  
GINLGNKDLKYMKRQMSNYKKTITOTSKVINDSDITILLTNDCTIRLSLVLYLN  
RHKKILLIDELIPIFLNLSNVHDELNSFLIGKASFPNIIIRNHPNVLIIISHVANT  
LSCDYYIVLRKGEITVRCSEYDVKTOSLSHLEMD"  
23896. .31533  
rRNA  
/gene="rRNA"  
/note="region containing small subunit, 5.8S and large  
subunit rRNA genes and spacer regions"  
23896. .31533  
gene  
/gene="rRNA"  
complement(join(31966. .32476,32675. .32775))  
/gene="MALP3.04"  
complement(join(31966. .32476,32675. .32775))  
/gene="MALP3.04"  
/note="MALP3.04, conserved hypothetical membrane protein,  
len: 203 aa, similarity: P. falciparum chromosome 2,  
PF80110W, 096126 predicted integral membrane protein (255  
aa), fasta scores: opt: 335, E(): 4.9e-15, (36.1% identity  
in 191 aa overlap)"  
/codon\_start=1  
/product="conserved hypothetical membrane protein,  
MALP3.04"  
/protein\_id="CAB63559.1"  
/db\_xref="GI:6594247"  
/translation="MKSKYTFINTYLLIFLLFLFYNYNDTSKTKFNNTIDIN  
RFRKIIABSEBOKYFWEDSCLTINEELTRPHEHSDPIPEHTENDKINELSTINS  
TKIMKETIKKRONYKETQDMNHNMRDPMHYWANTYLYKVKLLINITLKLDTLPI  
HDKETITTYKIQIEDIEYFLFMQVEMRLITLLEFYKNKE"  
complement(32477. .32486)  
/gene="MALP3.04"  
/note="potential splice acceptor sequence"  
complement(32669. .32674)  
/gene="MALP3.04"  
/note="potential splice donor sequence, aag/gtatat"  
join(36657. .36743,36864. .37343)  
/gene="MALP3.05"  
join(36657. .36743,36864. .37343)  
/gene="MALP3.05"  
/note="MALP3.05, hypothetical protein, len: 188 aa"  
/codon\_start=1  
/product="hypothetical protein, MALP3.05"  
/protein\_id="CAB63560.1"  
/db\_xref="GI:6594248"  
/translation="MRIKMSGIFPILICISYICVEPCNKCMISYRKLLMYSEN  
CFNYSIDRSIAEGSESEKTKVDIPNIELKSLNINTEEEKETVGSFMDNNLN  
IANEVLKNIHSEFTIENIFSLINDSSKSPVLTPLKEGSLFPHMLNVPKLLDLCO  
RNPPLHILGLIVILAAIYVFNFKNPEC"  
36744. .36749  
misc\_feature  
/gene="MALP3.05"  
/note="potential splice donor sequence, aag/gtatga"  
36854. .36863  
misc\_feature  
/gene="MALP3.05"  
/note="potential splice acceptor sequence"  
complement(join(38049. .39995,40210. .40284))  
/gene="garp"  
complement(join(38049. .39995,40210. .40284))  
/gene="garp"  
/note="MALP3.06, garp, len: 673 aa, similarity: almost  
identical to GARP PLAF (678 aa), fasta scores: 97.68  
identity in 678 aa overlap"  
/codon\_start=1  
/product="hypothetical garp protein"  
/protein\_id="CAB63561.1"  
/db\_xref="GI:6594249"  
/translation="MNVLFYSNICILFFVVCITLNFSTKCFNSGLKNONILNKSPDS  
ITGRLNETELEKKNKDNKSELTLEKDEKDDVPTTSNDNLKNNNNNEISSDTP  
TIIINVDKDNENSVYDKKDKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKK  
DKKKESEVMSLYTKQHKPKNAIEGEBELDEWSEINNAOGGLSSPYOYE  
OGGCGIISVHEHSNDPKDKNISDEKEDHOEMLKTLDKKKKKKKKKKKKK  
KIEKKKKOEKKEKK

HDKENEETMQPDOTSEETNNEIMVPLPSPLTDVTPPEHKEGEGHEKHEGHEK  
EHKEHEHEHEKHEKHEKSKGKKDKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKK  
DKDVEELINLEDKACEPOHKTIVSRKSPGKOLJIDPELOLTKDSSKVEKNLSTO  
EOLGICTIRVAVVRRRDNHKKKKKKKIEBAEOLKKRHYDKEDKKESEVESEKVO  
EDEEEVEDE  
DDDAEEDDDDAEEDDDDEDEDEDEDEDEDEDEDEDEDEDEDEDEDEDEDEDE  
complement(39996. .40005)  
/gene="garp"  
/note="potential splice acceptor sequence"  
complement(40204. .40209)  
/gene="garp"  
/note="potential splice donor sequence, aag/gtaaca"  
join(45401. .46396,46562. .50233)  
/gene="MALP3.07"  
misc\_feature  
Query Match 6.8% Score 80.2; DB 3; Length 67970;  
Best Local Similarity 47.7%; Pred. No. 0.00022;  
Matches 335; Conservative 0; Mismatches 358; Indels 10; Gaps 3;  
OY 50 atgaacacaaagaagaagtagtgcaccccttatatatatatatgcat 109  
DB 8027 ATTAATTAATTAATTAAGAAATATTATCTTTGTTATTAATTAATTAATG 8086  
OY 110 gcatgagaccatgctatgtaaggttaagatgagtgatgagatgca 169  
DB 8087 GAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAC 8143  
OY 170 gcaatgagtttttctgtgtggtatccatgagatgagcgaatattatatat 229  
DB 8144 AAT 8203  
OY 230 aatgataatagatgattatctctgtaatttgcgaatagatgataaacagctcaat 289  
DB 8204 ATTAATAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 8263  
OY 290 gtgagtgaccagtgatgcaaatgaccactgcagctgggagcagtgatgttcaatcac 349  
DB 8264 ATTAAT 8323  
OY 350 aactcaattgaaactaaactaaactaaactgattgattaaactttagtga---- 405  
DB 8324 TAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 8383  
OY 406 attcaccggtgccaatcaatcaatcaatcaatcaatcaatcaatcaatcaatcaatcaat 465  
DB 8384 ATTAATAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 8443  
OY 466 ttaaaatttattgatttgaatgaactcaat---tacatcaaaaaaactcaatcaat 522  
DB 8444 ATAAATATATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 8503  
OY 523 taatctctatgtatataatattgaataataataatgaatgaacttaacttaactcgcggttc 582  
DB 8504 TAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 8563  
OY 583 tcttaataaaaaacagtataatctgagctagattcaacagctatcaacagtcgacag 642  
DB 8564 ATTAATAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 8623  
OY 643 acaattataaattataatattatttttctataaagaagcactcctaattgttaaa 702  
DB 8624 ATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 8683  
OY 703 tatagtctaaacataataaattatttctgtatctt 745  
DB 8684 TATATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 8726  
RESULT 9  
AL513330 202645 bp DNA linear HTG 10-JUL-2001  
LOCUS AL513330  
DEFINITION Homo sapiens chromosome 1 clone RP11-469E8, \*\*\* SEQUENCING IN  
PROGRESS \*\*\*, 7 unordered pieces.

```

ACCESSION      AL513330
VERSION        AL513330.9 GI:13990571
KEYWORDS       HTG; HTGS_PHASE1; HTGS_CANCELLED.
SOURCE         human.
ORGANISM       Homo sapiens
REFERENCE      Mammalia; Euthetia; Primates; Catarrhini; Homnidae; Homo.
AUTHORS        McIay,K.
TITLE          Direct Submission
JOURNAL        Submitted (09-JUL-2001) Sanger Centre, Hinxton, Cambridgeshire,
               CB10 1SA, UK. E-mail enquiries: humquere@sanger.ac.uk Clone
               requests: clonerequests@sanger.ac.uk
               On May 7, 2001 this sequence version replaced gi:13396737.
COMMENT        -----
               Genome Center
               Center: Sanger Centre
               Center code: SC
               Web site: http://www.sanger.ac.uk
               Contact: humquere@sanger.ac.uk
               -----
               Project Information
               Center project name: BA469E8
               -----
               Summary Statistics
               Sequencing program: XGAP4; version 4.5
               Sequencing vector: plasmid: l08752; 100% of reads Chemistry:
               Chemistry: dye-terminator ET-amersham; 0% of reads Chemistry:
               Dye-terminator Big Dye; 99% of reads
               Consensus quality: 200655 bases at least Q40
               Consensus quality: 201365 bases at least Q30
               Consensus quality: 201775 bases at least Q20
               Insert size: 202045; sum-of-contigs
               Insert size: 161399; 19.6% error; agarose-fp
               Quality coverage: 6.15x in Q20 bases; sum-of-contigs Quality
               coverage: 7.70x in Q20 bases; agarose-fp
               -----
               * NOTE: This is a 'working draft' sequence. It currently
               * consists of 7 contigs. The true order of the pieces
               * is not known and their order in this sequence record is
               * arbitrary. Gaps between the contigs are represented as
               * runs of N, but the exact sizes of the gaps are unknown.
               * This record will be updated with the finished sequence
               * as soon as it is available and the accession number will
               * be preserved.
               *
               *   1 . . 7983: contig of 7983 bp in length
               *     7984 8083: gap of 100 bp
               *     8084 16795: contig of 8712 bp in length
               *     16796 16895: gap of 100 bp
               *     16896 73805: contig of 56910 bp in length
               *     73806 73905: gap of 100 bp
               *     73906 147587: contig of 73662 bp in length
               *     147588 147687: gap of 100 bp
               *     147688 178901: contig of 31214 bp in length
               *     178902 179001: gap of 100 bp
               *     179002 181519: contig of 2516 bp in length
               *     181520 181619: gap of 100 bp
               *     181620 202645: contig of 21026 bp in length.
               Location/Qualifiers
                   1 . . 202645
                   /organism="Homo sapiens"
                   /db_xref="taxon:9606"
                   /chromosome="1"
                   /clone="RP11-469E8"
                   /clone_id="RPC1-11.2"
                   /clone_lib="RPC1-11.2"
                   1 . . 7983
                   /note="assembly fragment: 00874
                   fragment_chain:1
                   clone_end:SP6
                   vector_side:left"
                   8084 . . 16795
                   /note="assembly fragment: 01844
                   fragment_chain:1"
                   16896 . 73805
                   /note="assembly fragment: 01698
                   fragment_chain:1"

```

	misc_feature	73906..147587	/note="assembly_fragment:03055 fragment_chain:1"
	misc_feature	147688..178901	/note="assembly_fragment:01073 fragment_chain:1"
	misc_feature	179002..181519	/note="assembly_fragment:01557 fragment_chain:1"
	misc_feature	181620..202645	/note="assembly_fragment:01958 fragment_chain:1 clone_end:r7 vector_side:right"
BASE COUNT	67892 a 34283 c 34403 g 65461 t	606 others	
ORIGIN			
Query Match	6.8%;	Score 80.2;	DB 2; Length 202645;
Best Local Similarity	49.9%;	Pred. No. 0.00017;	
Matches 235; Conservative	0;	Mismatches 228;	Indels 8; Gaps 1
Oy	336 attttccaatcacaccccaatttgaaacaataaataaagaatttcgatatctaact	395	
Db 202524	AATATTAAATATATTAATTTTAAATTTTAAATATATATTTTAAATATTTTAAATATAA	202465	
Oy	396 tattggtaattcccggttgcctaatcatattatattaataaacgatagiatctt	455	
Db 202464	TATATATTTTTAAATATATATTTAAATATATATATATTTTAAATATATATTTTATAT	202405	
Oy	456 gataattcaataaaatttatltggaattgaaigaacctcaattacacacaaaaacct	515	
Db 202404	AAATATATATTTTAAATTTATATACATTTTAAATATATTTTAAATATTTTATATCAAAAATATTTTATA	202345	
Oy	516 atcaaatcaatactcctaigtgatacaatttgaataataaagatbaecctta-----	569	
Db 202344	TAAAAATTTTATTTTATTTTAAAAAGATATATTTTAAATTTTAAAAATATATATTTT	202285	
Oy	570 --aatctcgagtctctctataaaaaaccgafataatttgscgtagatttaacgcatt	627	
Db 202284	ATAATATATTAATTTTATTTATATATGTCATCATATATATATTTATATATATATATATATTA	202225	
Oy	628 ttcaaacttgcgcaggacaactatlaaanaataaattatatttttctcaataaagcact	687	
Db 202224	TACTATATATTTATTTATTTAT	202165	
Oy	688 tccctaactgttaaaataatagctcctaacaactaataaataattcatttgcattccttgg	747	
Db 202164	TATATATGATATATATATATATTTATATTTATATATATATATATATATATATATATATTA	202105	
Oy	748 cagtagtgtagaggggtgcgcacaataaatttgctcctaataataatlgat	798	
Db 202104	TATATTTATTTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATCAT	202054	
RESULT 10			
LOCUS	AC096550	181486 bp	DNA linear HTG-18-SEP-2001
DEFINITION	Homo sapiens chromosome 2 clone RP11-20P13, WORKING DRAFT SEQUENCE.		
ACCESSION	AC096550 AC013666		
VERSION	AC096550.1 GI:15638715		
KEYWORDS	HTG: HTGS_PHASE1; HTGS_DRAFT.		
SOURCE	human.		
ORGANISM	Homo sapiens		
AUTHORS	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.		
JOURNAL	The sequence of Homo sapiens clone		
REFERENCE	Unpublished		
REFERENCE	2 (bases 1 to 181486)		
AUTHORS	Waterston,R.H.		





```

source
1. .14867
/organism="Plasmodium falciparum"
/db_xref="taxon:5833"
/chromosome="2"
/complement(1570. .2424)
/gene="PF0490c"
/complement(1570. .2424)
/gene="PF0490c"
/codon.start=1
/note="predicted by GlimmerM"
/product="hypothetical protein"
/protein_id="AAC71887.1"
/db_xref="GI:3845198"

gene
11241. .14606
/gene="PF0495w"
11241. .14606
/gene="PF0495w"
/note="predicted by GlimmerM"
/codon.start=1
/product="hypothetical protein"
/protein_id="AAC71888.1"
/db_xref="GI:3845198"

CDS
/translaton="MFAKKNIVREGVNICFSTYKLLYKSNFVYVNLNENR
TNKTKKSAQSLFDGKLIHKLILFKMLPKYCKKCYCISAKERYKLLDEKRC
FNYISLDDIIOVAKIFDELDTKTDYNYEIVKNIDKNVKNLNEYKKNLDTFHR
EILGICCNKIMSYIHEMGNELIHLIFPRMKNNDNLILFYNYNYEDHLYLN
HEIYKLEIFNKYLNNSNIPFNKNLIOEMFNLYPREIKENYIKNNKRYTKL
CPKFEHNVNDHIDNEKILNLILYVNSIIDIOMNKNLNNLNNLNNLNNLNNL
NEFTLKKKRYNDMTYKLEVIKTHHILCDKTKNLFESDIDYTLNLSLNK
FLNKITDKNFILYFECILILNINFAVNSQISLISKNYINLNNVYIVNV
LFNDIMKFSLYCNIEFKRIKTEENAVLIHNDQTNSENKIMIDIIOKRIKY
IIFYMENYKDFEFLKSDLSIKLSITFEVKINEVNSYDFYLFNNISCIILN
NRSVKKYKDTYIYLNDSEVYKIKNDRTKKKNEFLSSMKELICKNLSVSN
RYIKLHEEDNPOKDOYVSLTFINLFPDKIIPHAYTNMCHYKRYNYPCML
INEDIISLLTTSKROYFIENNSNDYCKRELHLKNTIIDILKNTLNTYSISD
NISKIFISLNSKYTCVNNLLESQSEFEVTKSKGGHMMNNLNDNNSE
KYEHRLEYKKNLFIENKILNINFLYKKNRYLYLKOELCINLENTL
KILLYANNLNYEMGYVCEMLERYSKSEONLEKYNNKVEHMEPKILCHIE
DDYIEMNTMYLVEYDKNINSEROSLNNSTNDRFLDEIKKKYKLNNTLILH
NNYKLNKSNNSNGNISNLKDDKNHNNEMDLIDNKNENKLOEKOONCNCN
CKRYLVNDIINIGFLKMEKKELFELYWYLCYTRFKRYVSSSLFHMDFKIIT
DMNLKYLCENYIKNECAFVYIDVIFKER"

BASE COUNT      6284 a 1019 c 1106 g 6458 t
ORIGIN

Query Match      6.7%: Score 78.4; DB 3: Length 14867;
Best Local Similarity 48.18; Pred. No. 0.00061;
Matches 260; Conservative 0; Mismatches 271; Indels 9; Gaps 1;

QY 210 aaatttatatatataatgaataatgatgtatctctgtatattgtgaa 269
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 8497 AATTAAGAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 8438

QY 270 atagaatcaaacagcctcaatgtgaggtgacaggttgcataatgaccactgacgtgagggc 329
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 8437 TAAATATATATATATATATATCAAGAAACAAATATATATATATATATATAT 8378

QY 330 atggtagttttcaatcacacacacatttgaataatcaataataaaagtttagatta 389
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 8377 TAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 8318

QY 390 ttaaatatagtagtaatccgggttgccaatcaatattatataataaacagatagt 449
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 8317 AAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 8258

QY 450 attttgataatataatcaaatattatgtatt-----tgaatgaacacaaatlac 500

```

```

DB 8257 ATATATCATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 8198
QY 501 atcacaanaaaccaatcaataatcatctatgtgataataattagaacataaagat 560
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 8197 AATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 8138
QY 561 taaccttaaatctcaggttctctctataaaaaaacagcataatggcctagattaca 620
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 8137 ATATATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 8078
QY 621 gctattatcaactgycgacacattataaataatataatatttcttcaata 680
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 8077 TATATATTTTATATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 8018
QY 661 aagccttcctcaattgttataaataatgtctcaaacacataataatattatgtgta 740
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 8017 TTTATATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 7958

```

```

RESULT 12
AC074035      189214 bp      DNA      linear      HTG 30-NOV-2001
LOCUS
DEFINITION
Homo sapiens chromosome Xp clone Rpl1-70P20, WORKING DRAFT
SEQUENCE, 2 unordered pieces.
AC074035
AC074035      21      GI:17149341
KEYWORDS
HTG: HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP; HTGS_ACTIVERPIN.
SOURCE
human.
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

```

## REFERENCE

## AUTHORS

1 (bases 1 to 189214)

Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C., Alsbrooks,S.L., Amaratunge,H.C., Are,J.R., Banks,T., Barbata,J., Benton,J., Bimage,K., Blankenburg,K., Bonnin,D., Bouck,d., Bowie,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P., Buhay,C., Butler,P., Burrell,C., Burrell,K.L., Byrd,N.C., Caron,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D., Chen,R., Chen,Z., Chowdhry,I., Christopoulos,C., Cleveland,C.D., Cox,C., Coyle,M.D., Dathorne,S.R., David,R., Davila,M.L., Davis,C., Day-Carroll,L., Dederich,D.A., Delaney,K.R., Delgado,O., Deun,A.L., Ding,Y., Dinh,H.H., Douthwaite,K.J., Draper,H., Dugan-Rocha,S., Duplin,K.J., Earhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Escotto,M., Falls,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P., Frantz,P., Gabisl,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R., Gorrell,J.H., Guevara,W., Gunatane,P., Hale,S., Hamilton,K., Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A., Hernandez,J., Hernandez,O., Hodgson,A., Hogues,M., Holloway,C., Hollins,B., Homs,F., Howard,S., Huber,J., Huylk,S., Hume,J., Jackson,L.E., Jacobson,B., Jia,Y., Johnson,R., Jolivet,S., Joudah,S., Karlsson,E., Kelly,S., Khan,U., King,L., Koryah,J., Kovar,C., Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C., Lewis,L., Li,J., Li,Z., Lichtarge,O., Iieu,C., Liu,J., Liu,W., Louisaged,H., Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R., Ma,J., Maheshwari,M., Mapua,P., Martin,R., Martindale,A., Martinez,E., Massey,E., Mawhney,E., Mcleod,M.P., Meador,M., Mel,G., Metzger,M., Miner,G., Miner,Z., Mitchell,P., Mohabbat,K., Morgan,M., Morris,S., Moser,M., Neal,D., Newton,J., Newton,N., Nguyen,A., Nguyen,N., Nickerson,E., Nwokweto,S., Ogulu,M., Okuwon,G., Ogunye,N., Oyiedo,R., Pace,A., Payton,B., Peery,J., Peters,L., Pickens,R., Primus,E., Pu,L.L., Quiles,M., Ren,Y., Rives,M., Rojas,A., Rojibokan,I., Rolfe,M., Ruiz,S., Savery,G., Scherer,S., Scott,G., Shen,H., Shooshtrai,N., Sisson,I., Sodergren,E., Sonatke,T., Sparks,A., Stanley,H., Stone,H., Sulton,A., Swalek,A., Taber,P., Tamerisa,A., Tamerise,K., Tang,H., Tansey,J., Taylor,C., Taylor,T., Tellford,B., Thomas,N., Thomas,S., Usmani,K., Vasquez,L., Vera,Y., Villalob,D., Vinson,R., Wall,R., Wang,S., Ward-Moore,S., Warren,R., Washington,C., Watlington,S., Williams,G., Williamson,A., Wlaczek,R., Wooden,S., Wolley,K., Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorilla,S., Nelson,D., Weinstock,G. and Gbbs,R.

Direct Submission

TITLE

2 (bases) 1 to 189214)  
 Worley, K.C.  
 Direct Submission  
 Submitted (12-Jul-2000) Human Genome Sequencing Center, Department  
 of Molecular and Human Genetics, Baylor College of Medicine, One  
 Baylor Plaza, Houston, TX 77030, USA  
 On Nov 29, 2001 this sequence version replaced gi:15327947.

---

Genome Center

Center: Baylor College of Medicine  
 Center code: BCM  
 Web site: <http://www.hgsc.bcm.tmc.edu/>  
 Contact: hgsc-help@bcm.tmc.edu

---

Project Information

Center project name: HRI1  
 Center clone name: RP11-707P20  
 Summary Statistics

---

Sequencing vector: Plasmid: M77789  
 Sequencing vector: M13: L08821

Chemistry: Dye-terminator Big Dye: 100% of reads  
 Assembly program: Phrap: version 0.990329  
 Consensus quality: 190931 bases at least Q40  
 Consensus quality: 191911 bases at least Q30  
 Consensus quality: 192818 bases at least Q20  
 Estimated insert size: 190344, sum-of-configs estimation  
 Quality coverage: 0x in Q20 bases; agarose-IP estimation  
 Quality coverage: 17.7x in Q20 bases; sum-of-configs estimation

```

* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/genbank\_draft\_data.html).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 2 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
*       1 130843: contig of 130843 bp in length
* 130844 130943: gap of unknown length
* 130944 189214: contig of 58271 bp in length.

```

FEATURES	source	Location/Qualifiers
		1..189314
		/organism="Homo sapiens"
		/db_xref="taxon:9606"
		/chromosome="Xp"
		/clone="RP11-707p20"
BASE COUNT	57784 a 35775 c 35259 g 60296 t	100 others
ORIGIN		

[illegible]

Db	54092	TTAATTAATAAATTATATATCAATATATTAATTAATTAATAAATTTATATATCAATATATTAATAATAA	54151
Qy	510	aacctcaacaaatcaatactctatgctacataalttagaataataaagattaaccttla	569
Db	54152	TAAAAATATATATCAATATATTAATTAATTAATAAATTTATATCAATATATTAATTAATAAATTTAT	54211
Qy	570	aatctcgagttctctctataaaaaaacacgctaatttggcgctagatttaacgactatct	629
Db	54212	ATATCAATATATTAATTAATTAATTAATAAATTAATATCAATATATTAATATATTAATATATTTAT	54271
Qy	630	caactgscgcggacaacttattataaataactatattcttctctaataaagaccttc	689
Db	54272	ATTATTAATATTAATTTTATATTAATTAATTAATTAATTTATATTTATTAATATGATATTAATAATATATTA	54331
Qy	690	ctaatctgtaaatatactatgcctaaacactaataaataatattatgttgcatacctt	745
Db	54332	TAAATATGATGATATATTTATATGTAATATGATATATTGAATTAATAATATACCATATTT	54387

RESULT 13

AC108683

LOCUS AC108683 208319 bp DNA linear HTG 31-JAN-2002

DEFINITION Homo sapiens chromosome Xp clone RP11-359J11, WORKING DRAFT

SEQUENCE, 3 unordered pieces.

ACCESSION AC108683

VERSION AC108683.1 GI:18449744

KEYWORDS HTG; HTGS\_PHASE1; HTGS\_DRAFT; HTGS\_FULLTOP.

SOURCE human

ORGANISM	REFERENCE
Homo sapiens	1 (bases 1 to 208319)
Eukaryota:Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.	
AUTHORS	
Muzny,D.M., Adams,C., Adio-Oduola,B., Alt-osman,F.R., Allen,C., Albrittcks,S.L., Amaratunge,H.C., Are,J.R., Ayele,M., Banks,T., Barbata,J., Benton,J., Blinaghe,K., Blankenburg,K., Bonini,D., Bouck,J., Bowle,S., Briviera,M., Brown,E., Brown,M., Bryant,N.P., Bunay,C., Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C., Carron,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Z., Chowdhry,I., Christopoulos,C., Cleveland,C.D., Cox,C., Coyle,M.D., Dathorne,S.R., David,R., Davila,M.L., Davis,C., Davy-Carroll,L., Dederich,D.A., Delaney,K.R., Delgado,O., Denn,A.L., Ding,Y., Dinl,H.H., Douthwaite,K.J., Draper,H., Dugan-Rocha,S., Dublin,K.J., Earhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Escotto,M., Falls,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P., Frantz,P., Gabisi,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R., Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S., Hamilton,K., Harris,C., Harris,K., Hart,M., Haylak,P., Hawes,A., Hernandez,J., Hernandez,O., Hodgson,A., Hogues,M., Holloway,C., Hollins,B., Homsi,F., Howard,S., Huber,J., Hultj,S., Hune,D., Jackson,L.E., Jacobson,B., Jia,Y., Johnson,R., Jolivet,S., Joudas,S., Kailsson,E., Kelly,S., Khan,U., King,L., Kovach,J., Kovar,C., Kratochvic,T., Kureshi,A., Landry,N., Leal,B., Lewis,L.C., Lewis,L., Li,J., Li,Z., Lichtenarge,O., Lieu,C., Liu,J., Liu,W., Louised,H., Lozodo,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R., Ma,J., Maheshwari,M., Mapua,P., Martin,R., Martindale,A., Martinez,E., Massey,E., Mawhney,E., Mcleod,M.P., Meador,M., Mel,G., Melzer,M., Miner,G., Miner,Z., Mitchell,T., Mohabbat,K., Morgan,M., Morris,S., Moser,M., Neal,D., Newton,J., Newton,N., Nguyen,A., Nguyen,N., Nguyen,N., Nickerson,E., Nwokekwu,S., Ogun,M., Okunoru,G., Orangunye,N., Oviedo,R., Pace,A., Payton,B., Peety,J., Perez,L., Peters,L., Plickens,R., Primus,E., Pu,L.L., Quiles,M., Ren,Y., Rivers,M., Rojas,A., Rojibokan,I., Rolfe,M., Ruiz,S., Savery,G., Scherer,S., Scott,G., Shen,H., Shooshbari,N., Sisson,I., Sodergren,E., Sonaike,T., Sparks,A., Stanley,H., Stone,H., Sulton,A., Svatek,A., Tabor,P., Tamerista,A., Tamerista,K., Tang,H., Tamey,J., Taylor,C., Taylor,T., Telitod,B., Thomas,N., Thomas,S., Usmani,K., Vasquez,L., Vera,V., Villalon,D., Vinsion,R., Wall,R., Wang,S., Ward-Moore,S., Warren,R., Washington,C., Wallington,S., Williams,G., Williamson,A., Wlaczek,R., Wooden,S., Worley,K., Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorilla,S., Nelson,D., Weinstein,G. and Gibbs,R.	
Direct Submission	

RESULT	14
PEMAL3P5	86827 bp DNA linear INV 04-MAY-2000
LOCUS	
DEFINITION	Plasmodium falciparum MAL3P5, complete sequence.
ACCESSION	AL034556 AL008971 AL008972 AL008978 AL010141 AL010153 AL010162 AL010206 AL010210 AL1391179
VERSION	AL034556.3 GI:7711064
KEYWORDS	HTG: centromerine; CTRP protein; initiation factor E4; Serine/threonine protein phosphatase. malaria parasite P. falciparum.
SOURCE	
ORGANISM	Plasmodium falciparum
REFERENCE	Plasmodium falciparum
AUTHORS	Enkaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium. 1 (bases 1 to 86827) Bowman,S., Lawson,D., Basham,D., Brown,D., Chillingworth,T., Churcher,C.M., Craig,A., Davies,R.M., Devlin,K., Feltwell,T., Gentles,S., Gilliam,R., Hamlin,N., Harris,D., Holroyd,S., Hornsby,T., Horrocks,P., Jagels,K., Jaissal,B., Kyes,S., McLean,J., Moule,S., Mungall,K., Murphy,L., Oliver,K., Quail,M.A., Rajandream,M.-A., Rutter,S., Skelton,J., Squares,R., Squares,S., Stulson,J.E., Whitehead,S., Woodward,J.R., Newbold,C. and Barrell,B.G. The complete nucleotide sequence of chromosome 3 of Plasmodium falciparum Nature 400 (6744), 532-538 (1999) 2 (bases 1 to 86827) Bowman,S., Skelton,J., Churcher,C., Lawson,D., Quail,M. and Barrell,B. Unpublished 3 (bases 1 to 86827) Lawson,D., Bowman,S. and Barrell,B. Direct Submission Submitted (17-DEC-1998) P. falciparum Genome Sequencing Consortium, The Sanger Centre, Wellcome Trust Genome Campus, Hinxton, Cambridge CB10 1SA, UK On May 14, 2001 this sequence version replaced gi:2982535 gi:2982536 gi:2694454 gi:2982554 gi:2982562 gi:2894409 gi:2982572 gi:2982574 gi:4493931. For more information about this sequence or the Malaria Project, see <a href="http://www.sanger.ac.uk/Projects/P_falciparum">http://www.sanger.ac.uk/Projects/P_falciparum</a> . Location/Qualifiers 1. .86827 /organism="Plasmodium falciparum" /strain="3D7" /db_xref="taxon:5833" /chromosome="3" /clone="MAL3P5" /join(332..603,826..1023,1197..1301,1458..2942) /gene="PFC0575w, MAL3P5.1" /join(332..603,826..1023,1197..1301,1458..2942) /gene="PFC0575w, MAL3P5.1" /note="(PFC0575w (MAL3P5.1), Hypothetical protein, len: 689 aa, possible signal sequence, revised: added new exon 2"
FEATURES	
source	
COMMENT	
JOURNAL	
REFERENCE	
AUTHORS	
TITLE	
JOURNAL	

[illegible]



















DJ	26-MAR-2002	(first entry)
XX	Human immune system associated gene SEQ ID NO: 298.	
XX	Human; immune system disease; cytosine methylation; antiasthmatic;	
KW	antiartherosclerotic; antianemic; cytostatic; nootropic;	
KW	neuroprotective; anti-HIV; anticonvulsant; ophthalmological;	
KW	antiinflammatory; antirheumatic; antididiabetic; antiparasitic;	
KW	antiinflammatory; cancer; eye disease; arteriosclerosis; anaemia;	
KW	acute myeloid leukaemia; Alzheimer's disease; AIDS; epilepsy;	
KW	neurofibromatosis; rheumatoid arthritis; psoriasis; bowel disease;	
KW	gene; ds.	
XX	Homo sapiens.	
OS	WO200200928-A2.	
PN	03-JAN-2002.	
XX	02-JUL-2001; 2001WO-EP07537.	
XX	30-JUN-2000; 2000DE-1032529.	
PR	01-SEP-2000; 2000DE-1043826.	
PR	(EPIG-) EPIGENOMICS AG.	
PA	Olek A, Piepenbrock C, Berlin K;	
XX	WPI: 2002-130909/17.	
DR	Nucleic acid comprising fragment of chemically modified gene, useful	
PT	for diagnosis and treatment of diseases associated with abnormal	
PT	cytosine methylation -	
PS	Claim 1; SEQ ID NO 298; 32pp + Sequence Listing; German.	
XX	The present invention provides a number of human immune system associated	
CC	genes which are modified by the methylation of cytosines. The sequences	
CC	can be used in the diagnosis and treatment of immune system disorders,	
CC	including eye diseases such as retinopathy, neovascular glaucoma and	
CC	macular degeneration, arteriosclerosis, anaemia, cancer, acute myeloid	
CC	leukemia, Alzheimer's disease, AIDS, epilepsy, neurofibromatosis,	
CC	rheumatoid arthritis, psoriasis and inflammatory/intestative bowel	
CC	diseases. The present sequence is a gene of the invention.	
XX	Sequence 6071 BP; 1973 A; 46 C; 1013 G; 3039 T; 0 other:	
SQ		
Query Match	5.9%; Score 69; DB 24; Length 6071;	
Best Local Similarity	48.4%; Pred. No. 0.0013;	
Matches 256; Conservative 0; Mismatches 265; Indels 8; Gaps 2;		
OY	348 acacactcaattggaactaaataaaaaagatttagatatctaattatagttaaat 407	
DB	651 AAAAAGTCATCTCAAAAAACAAACAAACAAAACCTTATTAAAAAATTATTCTTTT 592	
OY	408 tcaagggttgcgtaaactaatatattaataaacgtagtatttttgataattaat 467	
DB	501 AAACCAAAAAAACCTTAAGTCATTTAACTTTTAAAAATAAAACTTTTAAACAAAAA 532	
OY	468 aaaatttatattgatctgaaatgaactcaattcac--acaaaanaactaatcaattaa 525	
DB	531 TTTCATTAANAATAATTTTAAATAAATTAATCATCTATCAAAAATAATTAANAATCTA 472	
OY	526 tatctatgtagataaatttgaaatatacaatgatgaaccttaaatcctggagtlctct 585	
DB	471 AACCTTATCTCTTAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATA 412	
OY	586 tataaaaaaccgataaatttgccagattgaagaagcatatattccaacatggccggaca 645	
DB	411 TATATAAATAATATATATCTTAAACAATAATTAAAAATAAATAAATAAATAAATAAATA 352	
OY	646 attataaataataaataattatatttttcctaataagcaactccctaattgttaaaatat 705	

D6	331	TTTATACCTTTTAAAAATATAA-----CTAATATAAACGTTTTATACCTTTTAAAAAAA	298
OY	706	atgctcaaacacataataaatttcattgtgtaaccttcggcagtagtgagtgct	765
Db	237	ACCTAATTAATTAACCTTTAATACCTATATATATATATTAACATTTTAAAAATACCAAT	238
OY	766	gacaataaattagtcgataaataataatgattggtggtctgtgaaagacagtgag	825
Db	237	AATTTCTAATAATATTTCTAACCACTTCATTTTCATTTTATCTAACAAATATAAAAAATATA	178
OY	826	gacaagccacctctcgaagtcaaaaggccattccaaaccaacca 874	
Db	177	TTTAAAAATTTATTTAATTAATTAACCAACCATTAATTTATTTACGCCAA	129
RESULT 9			
ASAS61076/c			
ID	ASAS61076	standard; DNA; 6071 BP.	
XX			
AC	ASAS61076;		
XX			
DT	29-JAN-2002	(first entry)	
XX			
DE	Human gene regulation-associated gene oligonucleotide #31.		
XX			
KW	Human; Gene regulation-associated gene; severe combined immunodeficiency;		
KW	cardiac damage; inflammatory response; Hemophilia; Werner syndrome;		
KW	asthma; HDR syndrome; congenital heart defect; Saethre-Chotzen syndrome;		
KW	renal disease; Preclempsia; cardiac allograft vascular disease;		
KW	colorectal cancer; thyroid cancer; oesophageal cancer; ds; tumour;		
KW	immunostimulant; cardiac; antiinflammatory; coagulant; antiasthmatic;		
KW	nephrotropic; gynecological; anti-tumour; immunosuppressive; cytostatic.		
XX			
OS	Homo sapiens.		
XX			
PN	WO200177375-A2.		
XX			
PD	18-OCT-2001.		
XX			
PF	06-APR-2001; 2001MO-EP03968.		
XX			
PR	06-APR-2000; 2000DE-1019058.		
PR	07-APR-2000; 2000DE-1019173.		
PR	30-JUN-2000; 2000DE-1032529.		
PR	01-SEP-2000; 2000DE-1043826.		
XX			
PA	(EPIG-) EPIGENOMICS AG.		
XX			
PI	Olek A, Piepenbrock C, Berlin K;		
XX			
DR	WPI: 2002-017470/02.		
XX			
PT	New nucleic acid sequences from chemically modified genes associated		
PT	with gene regulation, useful for analysing cytosine methylations for		
PT	diagnosis and therapy of diseases e.g. severe combined immunodeficiency		
PT	disease		
XX			
PS	Claim 1; SEQ ID No 32; 26pp; English.		
XX			
CC	The invention relates to 224 nucleic acid sequences comprising at least		
CC	18 bases of a chemically pretreated gene associated with gene regulation		
CC	selected from 43 known genes (or complementary sequences). The		
CC	chemical pretreatment converts cytosine bases unmethylated at the		
CC	5-position to uracil or another base with hybridisation behaviour		
CC	dissimilar to cytosine, to enable analysis of cytosine methylations.		
CC	The DNA sequences, oligomers (or sets/arrays) and method are		
CC	useful in the diagnosis of diseases (or predisposition to diseases)		
CC	associated with gene regulation and in therapy of such diseases, by		
CC	enabling analysis of the cytosine methylation patterns of such genes,		
CC	kits are provided. They are especially useful in diagnosis		
CC	and therapy of e.g. severe combined immunodeficiency disease, cardiac		
CC	disorders, haemophilia, solid tumours and cancer, Werner syndrome,		



PR	02-OCT-2000;	20000US-0236802.
PR	02-OCT-2000;	20000US-0237037.
PR	02-OCT-2000;	20000US-0237038.
PR	02-OCT-2000;	20000US-0237039.
PR	02-OCT-2000;	20000US-0237040.
PR	13-OCT-2000;	20000US-0239935.
PR	13-OCT-2000;	20000US-0239937.
PR	20-OCT-2000;	20000US-0240960.
PR	20-OCT-2000;	20000US-0241221.
PR	20-OCT-2000;	20000US-0241785.
PR	20-OCT-2000;	20000US-0241786.
PR	20-OCT-2000;	20000US-0241787.
PR	20-OCT-2000;	20000US-0241808.
PR	20-OCT-2000;	20000US-0241809.
PR	20-OCT-2000;	20000US-0241826.
PR	01-NOV-2000;	20000US-0246477.
PR	08-NOV-2000;	20000US-0246474.
PR	08-NOV-2000;	20000US-0246475.
PR	08-NOV-2000;	20000US-0246476.
PR	08-NOV-2000;	20000US-0246477.
PR	08-NOV-2000;	20000US-0246478.
PR	08-NOV-2000;	20000US-0246523.
PR	08-NOV-2000;	20000US-0246524.
PR	08-NOV-2000;	20000US-0246525.
PR	08-NOV-2000;	20000US-0246526.
PR	08-NOV-2000;	20000US-0246527.
PR	08-NOV-2000;	20000US-0246528.
PR	08-NOV-2000;	20000US-0246532.
PR	08-NOV-2000;	20000US-0246509.
PR	08-NOV-2000;	20000US-0246610.
PR	08-NOV-2000;	20000US-0246611.
PR	08-NOV-2000;	20000US-0246613.
PR	17-NOV-2000;	20000US-0249207.
PR	17-NOV-2000;	20000US-0249208.
PR	17-NOV-2000;	20000US-0249209.
PR	17-NOV-2000;	20000US-0249210.
PR	17-NOV-2000;	20000US-0249211.
PR	17-NOV-2000;	20000US-0249212.
PR	17-NOV-2000;	20000US-0249213.
PR	17-NOV-2000;	20000US-0249214.
PR	17-NOV-2000;	20000US-0249215.
PR	17-NOV-2000;	20000US-0249216.
PR	17-NOV-2000;	20000US-0249217.
PR	17-NOV-2000;	20000US-0249218.
PR	17-NOV-2000;	20000US-0249244.
PR	17-NOV-2000;	20000US-0249245.
PR	17-NOV-2000;	20000US-0249246.
PR	17-NOV-2000;	20000US-0249265.
PR	17-NOV-2000;	20000US-0249265.
PR	17-NOV-2000;	20000US-0249297.
PR	17-NOV-2000;	20000US-0249299.
PR	17-NOV-2000;	20000US-0249300.
PR	01-DEC-2000;	20000US-0250160.
PR	01-DEC-2000;	20000US-0250391.
PR	05-DEC-2000;	20000US-0251030.
PR	05-DEC-2000;	20000US-0251988.
PR	05-DEC-2000;	20000US-0256719.
PR	06-DEC-2000;	20000US-0251479.
PR	08-DEC-2000;	20000US-0251855.
PR	08-DEC-2000;	20000US-0251866.
PR	08-DEC-2000;	20000US-0251869.
PR	08-DEC-2000;	20000US-0251989.
PR	08-DEC-2000;	20000US-0251990.
PR	11-DEC-2000;	20000US-0254097.
PR	05-JAN-2001;	20010US-0255678.
XX		
PA	(HUMA-) HUMAN GENOME SCI INC.	
XX		
PI	Rosen CA, Barash SC, Ruben SM;	
XX		
DR	WPI: 2001-483426/52.	
XX		
PT	Nucleic acids encoding human immune/hematopoietic antigen polypeptides	
PR	useful for preventing, diagnosing and/or treating cancers and	

PT metastasis -  
XX  
PS Disclosure; SEQ ID NO 37522; 3071bp + Sequence Listing; English.  
XX  
CC AAK54951 to AAK64702 encode the human immune/haematopoietic antigen (I)  
CC amino acid sequences given in AAM821170 to AAM91921. (I) have cytostatic  
CC activity, and can be used in gene therapy and vaccine production. (I)  
CC proteins and polynucleotides may be used in the prevention, diagnosis and  
CC treatment of diseases associated with inappropriate (I) expression. For  
CC example, they may be used to treat disorders associated with decreased  
CC expression by rectifying mutations or deletions in a patient's genome  
CC that affect the activity of (I) by expressing inactive proteins or to  
CC supplement the patients own production of (I). Additionally, (I)  
CC polynucleotides may be used to produce the secreted (I), by inserting  
CC the nucleic acids into a host cell and culturing the cell to express the  
CC protein. (I) proteins and polynucleotides may be used to prevent,  
CC diagnose and treat metastases of haematopoietic-related diseases, especially  
CC cancers and cancer metastases of haematopoietic-derived cells. AAK64703  
CC to AAK87594 represent human immune/haematopoietic antigen genomic  
CC sequences from the present invention. AAK54942 to AAK54950 and AAM82159  
CC represent sequences used in the exemplification of the present invention.  
XX  
XX Sequence 21313 BP; 5220 A; 5786 C; 5273 G; 5034 T; 0 other;

Query Match	5.9%	Score 68.8	DB 22	Length 21313
Best Local Similarity	48.2%	Pred. No. 0.0015		
Matches 193	Conservative 0	Mismatches 207	Indels 0	Gaps 0
QY 400	aggttaattcaagggttgccatcaatattatattaaataagcagtagtttggata	459		
DB 17149	AGGGCTATTATGGGGCTTATAGCTTAGGGACTGTGGATGAAACCGATATATATTATTA	17090		
QY 460	atttaattaaatttlaattggaattggaalgaactcaattacacacaaacctaatca	519		
DB 17089	TATTTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT	17030		
QY 520	aactaatatctatgtgataataattgagaataataagatgaaccttaactcgagt	579		
DB 17029	AT	16970		
QY 580	ttccctcttaaaaaacagataacttggcgctagatttaacagctattctcaactggcc	639		
DB 16969	AT	16910		
QY 640	aggacaattatbaaaatlaattatatttttcttaataaagcacttccaaattgcta	699		
DB 16909	TAT	16850		
QY 700	aaatatagtccaaacccctcaataaataattatttggtaacttggcagtgagtgga	759		
DB 16849	TAT	16790		
QY 760	ggtgcgtgacaataaattagtcataaataataatgagatt	799		
DB 16789	TAT	16750		
RESULT 11				
ABL33451/C				
ID ABL33451	standard; DNA; 19787 BP.			
XX ABL33451;				
XX				
DT 26-MAR-2002	(first entry)			
XX				
DE Human immune system associated gene SEQ ID NO: 1424.				
XX				
KM Human: immune system disease; cytosine methylation; antiasthmatic;				
KM antiatherosclerotic; antinaemic; cytosatic; nootropic;				
KM neuroprotective; anti-HIV; anticonvulsant; ophthalmologic;				
KM antineumatic; antiarthritic; antidiabetic; antiporiatic;				
KM antiinflammatory; cancer; eye disease; arteriosclerosis; anaemia;				

XX	acute myeloid leukaemia; Alzheimer's disease; AIDS; epilepsy;
KW	neurofibromatosis; rheumatoid arthritis; psoriasis; bowel disease;
KW	gene; ds.
XX	
OS	Homo sapiens.
XX	
PN	WO200200928-A2.
XX	
PD	03-JAN-2002.
XX	
PF	02-JUL-2001; 2001WO-EP07537.
XX	
PR	30-JUN-2000; 2000DE-1032529.
PR	01-SEP-2000; 2000DE-1045826.
XX	
PA	(EPIG-) EPIGENOMICS AG.
XX	
PI	Olek A, Piepenbrock C, Berlin K;
XX	
DR	WPI; 2002-130909/17.
XX	
PT	Nucleic acid comprising fragment of chemically modified gene, useful
PT	for diagnosis and treatment of diseases associated with abnormal
PT	cytosine methylation -
XX	
PS	Claim 1; SEQ ID NO 1424; 32pp + Sequence Listing; German.
XX	
CC	The present invention provides a number of human immune system associated
CC	genes which are modified by the methylation of cytosines. The sequences
CC	can be used in the diagnosis and treatment of immune system disorders,
CC	including eye diseases such as retinopathy, neovascular glaucoma and
CC	macular degeneration, arteriosclerosis, anaemia, cancer, acute myeloid
CC	leukaemia, Alzheimer's disease, AIDS, epilepsy, neurofibromatosis,
CC	rheumatoid arthritis, psoriasis and inflammatory/ulcerative bowel
CC	diseases. The present sequence is a gene of the invention.
XX	
SO	Sequence 19787 BP; 6302 A; 270 C; 3756 G; 9459 T; 0 other;

Query Match	5.9%	Score 68.6	DB 24	Length 19787
Best Local Similarity	51.9%	Pred. No. 0.0016		
Matches 203	Conservative	0	Mismatches 184	Indels 4
				Gaps 2
QY	336	attttcaaatcacacaactcaatttgaaacctaaattaaagaattagattat	395	
Db	2430	atttttaacacacaaaactatttttactaaattaaaaattcAAAAATT	2431	
QY	336	tattagtgtaattccagggctggcctaactatbattaaattaaacagat	455	
Db	2430	tctatttttAAATTTAACTTCACGTAATAAAATTAAGTTAAATTCAT	2371	
QY	456	gataatttaattaaatttatttgattgtaagtcaactcaactcaaaaa	515	
Db	2370	acttacaattatttAAAAATCATATAAACTAAATTAATTAAGCTTTT	2311	
QY	516	atcaaatiaaatacttaigtatataatttgaaatataaactgattaac	575	
Db	2310	ccAAATATACGATTTTGGTATTTAAATTAATAAACGATTA-TAAGT	2252	
QY	576	gaattctctataaaaaaacgctataattggcctagattcaacgctatt	635	
Db	2251	ctATTTTTCATTTATTAATTCACCTTAAAAATAACCTA--TTCCTT	2195	
QY	636	ggcgcaggaactatttaaataaactatatttttcttaataagaactct	695	
Db	2134	ttCATTTATCAATAAATAAATAATATTTTATTATTATTTTCTCAAA	2135	
QY	696	gttaaatatattgctcaaacactaaataaa 726		
Db	2134	tATTACCATACTAAATATCAATCTTTAA 2104		

ABL34175/C	standard; DNA; 113515 BP.
ABL34175	
ABJ34175;	
26-MAR-2002	(first entry)
Human immune system associated gene	SEQ ID NO: 2148.
Human; immune system disease; cytosine methylation; antiasthmatic; antiatherosclerotic; anti-naemic; cytosolic; nootropic; neuroprotective; anti-HIV; anticonvulsant; ophthalmological; antiinflammatory; antiarthritic; antidiabetic; antipsoriatic; antihematoma; cancer; eye disease; arteriosclerosis; anaemia; acute myeloid leukaemia; Alzheimer's disease; AIDS; epilepsy; neurofibromatosis; rheumatoid arthritis; psoriasis; bowel disease; gene; ds.	
Homo sapiens.	
WO200200928-A2.	
03-JAN-2002.	
02-JUL-2001; 2001WO-EP07537.	
30-JUN-2000; 2000DE-1032529.	
01-SEP-2000; 2000DE-1043826.	
(EPIG-) EPIGENOMICS AG.	
Olek A, Piepenbrock C, Berlin K;	
WPI: 2002-130909/17.	
Nucleic acid comprising fragment of chemically modified gene, useful for diagnosis and treatment of diseases associated with abnormal cytosine methylation	
Claim 1; SEQ ID NO 2148; 32pp + Sequence Listing; German.	
The present invention provides a number of human immune system associated genes which are modified by the methylation of cytosines. The sequences can be used in the diagnosis and treatment of immune system disorders, including eye diseases such as retinopathy, neovascular glaucoma and macular degeneration, arteriosclerosis, anaemia, cancer, acute myeloid leukaemia, Alzheimer's disease, AIDS, epilepsy, neurofibromatosis, rheumatoid arthritis, psoriasis and inflammatory/vulcerative bowel diseases. The present sequence is a gene of the invention.	
Sequence 113515 BP; 35172 A; 1174 C; 22520 G; 54649 T; 0 other;	
Query Match	5.9%; Score 68.6; DB 24; Length 113515;
Best Local Similarity	48.8%; Pred. No. 0.0017;
Matches 185; Conservative	0; Mismatches 194; Indels 0; Gaps 0
210	aaaatttatatatataataagaaataatgatgatattctctgtaatttggaa 269
25148	AAAAAATAAATAATATATATATCAAAACCTACTATTATTAATACATATATAAACCTTTAA 25089
270	atagattaaacagcccaatgtaggtgacccagttgcaaatgaccactgcactgg99c 329
25088	AAATTCATATAAAACATTATTAAATATATATTTACTATTATTAATATAATATTTCATA 25029
330	atgtagtatttccaacacccaactcaattgaaactaaataaataaagaattgatta 389
25028	ATATTACTTTTATTAACCAATTAATCCAAATATAAATAAATTTAAACCAAAACAATT 24969
390	ttaaatattagtgtaattccacggggttgctaatcaattatattatataaataacgattg 449
24968	TTTAACTCCCACTTAATAATTAACAATTTTAAATATATATATATTTAACTTAATT 24909

PS	Claim 1, SEQ ID NO 2148; 32pp + Sequence Listing; German.
XX	
CC	The present invention provides a number of human immune system associated
CC	genes which are modified by the methylation of cytosines. The sequences
CC	can be used in the diagnosis and treatment of immune system disorders,
CC	including eye diseases such as retinopathy, neovascular glaucoma and
CC	macular degeneration, arteriosclerosis, anaemia, cancer, acute myeloid
CC	leukaemia, Alzheimer's disease, AIDS, epilepsy, neurofibromatosis,
CC	rheumatoid arthritis, psoriasis and inflammatory/ulcerative bowel
CC	diseases. The present sequence is a gene of the invention.
XX	
SO	Sequence 113515 BP; 35172 A; 1174 C; 22520 G; 54649 T; 0 other;
	Query Match 5.9%; Score 68.6; DB 24; Length 113515;
	Best Local Similarity 48.8%; Pred. No. 0.0017;
	Matches 185; Conservative 0; Mismatches 194; Indels 0; Gaps 0
QY	210 aaaaatttatatatataataagaataatgatgatattctcgttaatttggaa 269
DB	25148 AAAAAATTAATAATATATATATCAAAACATCAATTTTAATACATATATATAACCTTTAA 25089
QY	270 atagattaaacagcgtcaatgtgaggtgacgaattgcaaatgacacgtcagacttgg 329
DB	25088 AAAATTCATTAATAAACATTTATTAATAATATATTCATTTTAATTAATAATTTTCATA 25029
QY	330 atggttgattttccaacatccaactcaatttgaagaacataaaagaattagatla 389
DB	25028 ATATTACCTTTTATTAACAATTAATCCAAATATATAAATAATTAACACAAACAATTT 24969
QY	390 ttaaatattataggttaattcaacgggtgttgctataatcatatcttaattaaacgatagt 449
DB	24968 TTAACCTCCCAACTTAATAATACAAACAATTTTAAATATATTCATTTATTTTAACCTAAATT 24909







PR 02-OCT-2000; 2000US-0237040.  
PR 13-OCT-2000; 2000US-0239935.  
PR 13-OCT-2000; 2000US-0239937.  
PR 20-OCT-2000; 2000US-0240960.  
PR 20-OCT-2000; 2000US-0241221.  
PR 20-OCT-2000; 2000US-0241785.  
PR 20-OCT-2000; 2000US-0241786.  
PR 20-OCT-2000; 2000US-0241787.  
PR 20-OCT-2000; 2000US-0241808.  
PR 20-OCT-2000; 2000US-0241809.  
PR 20-OCT-2000; 2000US-0241826.  
PR 01-NOV-2000; 2000US-0244617.  
PR 08-NOV-2000; 2000US-0246474.  
PR 08-NOV-2000; 2000US-0246475.  
PR 08-NOV-2000; 2000US-0246476.  
PR 08-NOV-2000; 2000US-0246477.  
PR 08-NOV-2000; 2000US-0246478.  
PR 08-NOV-2000; 2000US-0246523.  
PR 08-NOV-2000; 2000US-0246524.  
PR 08-NOV-2000; 2000US-0246525.  
PR 08-NOV-2000; 2000US-0246526.  
PR 08-NOV-2000; 2000US-0246527.  
PR 08-NOV-2000; 2000US-0246528.  
PR 08-NOV-2000; 2000US-0246532.  
PR 08-NOV-2000; 2000US-0246609.  
PR 08-NOV-2000; 2000US-0246610.  
PR 08-NOV-2000; 2000US-0246611.  
PR 08-NOV-2000; 2000US-0246613.  
PR 17-NOV-2000; 2000US-0249207.  
PR 17-NOV-2000; 2000US-0249208.  
PR 17-NOV-2000; 2000US-0249209.  
PR 17-NOV-2000; 2000US-0249210.  
PR 17-NOV-2000; 2000US-0249211.  
PR 17-NOV-2000; 2000US-0249212.  
PR 17-NOV-2000; 2000US-0249213.  
PR 17-NOV-2000; 2000US-0249214.  
PR 17-NOV-2000; 2000US-0249215.  
PR 17-NOV-2000; 2000US-0249216.  
PR 17-NOV-2000; 2000US-0249217.  
PR 17-NOV-2000; 2000US-0249218.  
PR 17-NOV-2000; 2000US-0249244.  
PR 17-NOV-2000; 2000US-0249245.  
PR 17-NOV-2000; 2000US-0249264.  
PR 17-NOV-2000; 2000US-0249265.  
PR 17-NOV-2000; 2000US-0249297.  
PR 17-NOV-2000; 2000US-0249299.  
PR 17-NOV-2000; 2000US-0249300.  
PR 01-DEC-2000; 2000US-0250160.  
PR 01-DEC-2000; 2000US-0250391.  
PR 05-DEC-2000; 2000US-0251030.  
PR 05-DEC-2000; 2000US-0251988.  
PR 05-DEC-2000; 2000US-0256719.  
PR 06-DEC-2000; 2000US-0251479.  
PR 08-DEC-2000; 2000US-0251856.  
PR 08-DEC-2000; 2000US-0251869.  
PR 08-DEC-2000; 2000US-0251989.  
PR 08-DEC-2000; 2000US-0251990.  
PR 11-DEC-2000; 2000US-0254097.  
PR 05-JAN-2001; 2001US-0259678.  
PA (HUMA-) HUMAN GENOME SCI INC.  
XX  
XX  
PI Rosen CA, Barash SC, Ruben SM;  
XX  
XX WPI: 2001-483426/52.  
XX  
XX Nucleic acids encoding human immune/hematopoietic antigen polypeptides,  
PT useful for preventing, diagnosing and/or treating cancers and  
XX metastasis -  
PS  
XX Disclosure: SEQ ID NO 41082; 3071pp + Sequence Listing; English.

CC AAK54951 to AAK64702 encode the human immune/haematopoietic antigen (I)  
CC amino acid sequences given in AAM82170 to AAM91921. (I) have cytosolic  
CC activity, and can be used in gene therapy and vaccine production. (I)  
CC proteins and polynucleotides may be used in the prevention, diagnosis and  
CC treatment of diseases associated with inappropriate (I) expression. For  
CC example, they may be used to treat disorders associated with decreased  
CC expression by rectifying mutations or deletions in a patient's genome  
CC that affect the activity of (I) by expressing inactive proteins or to  
CC supplement the patient's own production of (I). Additionally, (I)  
CC polynucleotides may be used to produce the secreted (I), by inserting  
CC the nucleic acids into a host cell and culturing the cell to express the  
CC protein. (I) proteins and polynucleotides may be used to prevent,  
CC diagnose and treat immune/haematopoietic-related diseases, especially  
CC cancers and cancer metastases of haematopoietic-derived cells. AAK64703  
CC to AAK87694 represent human immune/haematopoietic antigen genomic  
CC sequences from the present invention. AAK54942 to AAK54950 and AAM82169  
CC represent sequences used in the exemplification of the present invention.  
XX  
SQ Sequence 9706 BP; 3114 A; 1643 C; 1682 G; 3267 T; 0 other;

Query Match 5.7%; Score 66.8; DB 22; Length 9706;  
Best local Similarity 50.3%; Fred. No. 0.0033;  
Matches 190; Conservative 0; Mismatches 187; Indels 1; Gaps 1;

OY 351 actcaattgaaactaaataaagaattagattatattattaggttatca 410  
DB 7540 aatatataataataataataataataataataataataataataataata 7599  
OY 411 cgggttgctaaactaatattatattaaacagatagatttttgataatttaaa 470  
DB 7600 tatattatat-ataatataataataataataataataataataataataata 7658  
OY 471 atttatggatttgaagaactcaattcacacaaaacatcatcaataatatt 530  
DB 7659 attataataataattataattatattatataataataataataataattatt 7718  
OY 531 tatgtatataatttagaataataatagttaacctttaaactcgagtttccttaa 590  
DB 7719 tatataataataagttataatttatattatataataataataataattatatt 7778  
OY 591 aaaacagtataattggtgctagatttaacagctattatcaaacctggcaggacaattat 650  
DB 7779 attataaattataattttatgtataataattatattatataataataataatt 7838  
OY 651 taaataataattatttttttcttaaaagcacttcccttaattgtaataatagtc 710  
DB 7839 taattataataattataattataataataataataataataataataattttt 7898  
OY 711 taacacctaataataaa 728  
DB 7899 catataataataataaa 7916

Search completed: July 30, 2002, 11:22:21  
Job time: 4595 sec

---

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 30, 2002, 09:06:51 ; Search time 1681.15 Seconds  
(without alignments)  
9409.292 Million cell updates/sec

Title: US-09-530-663B-5  
Perfect score: 1172  
Sequence: 1 tftagagattgttgatgag.....agccgcgaatgacgcagaca 1172

Scoring table: IDENTITY NUC  
Gapop 10.0 , Gapext 1.0

Searched: 13736207 seqs, 6748477542 residues  
Total number of hits satisfying chosen parameters: 27472414

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : EST:  
1: em\_estbta:\*  
2: em\_estlhm:\*  
3: em\_estin:\*  
4: em\_estmu:\*  
5: em\_estov:\*  
6: em\_estcpl:\*  
7: em\_estcro:\*  
8: em\_hlc:\*  
9: gb\_est1:\*  
10: gb\_est2:\*  
11: gb\_hlc:\*  
12: gb\_gss:\*  
13: em\_gss\_hum:\*  
14: em\_gss\_inv:\*  
15: em\_gss\_pln:\*  
16: em\_gss\_vrt:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	99	8.4	1101	12	CNS0039G
2	94	8.0	1101	12	CNS00EVL
3	93.4	8.0	1101	12	CNS00EVL
4	91.4	7.8	1101	12	CNS00EVL
5	89.2	7.6	1101	12	CNS00EVL
6	84.8	7.2	1101	12	CNS00EVL
7	84.8	7.2	1101	12	CNS00EVL
8	84	7.2	1101	12	CNS00EVL
9	82.2	7.0	1101	12	CNS00EVL
10	81.2	6.9	1101	12	CNS00EVL
11	80.6	6.8	1101	12	CNS00EVL
12	79.6	6.9	1101	12	CNS00EVL
13	79	6.7	1101	12	CNS00EVL
14	78.4	6.7	1101	12	CNS00EVL
15	78	6.7	1101	12	CNS00EVL
16	77.6	6.6	1101	12	CNS00EVL

18	77	6.6	1101	12	CNS00EVL	AL069526 Drosophila
19	77	6.6	1143	9	AL565457	AL565457 Drosophila
20	77	6.6	1143	9	AL565457	AL565457 Drosophila
21	76.2	6.5	1101	12	CNS00EVL	AL060732 Drosophila
22	76	6.5	1225	12	CNS00EVL	AL061721 Drosophila
23	75.8	6.5	1092	12	CNS00EVL	AL175696 Tetradon
24	75.8	6.5	1101	12	CNS00EVL	AL057419 Drosophila
25	75.6	6.5	886	12	BH177277	BH177277 008.L.22-
26	75.6	6.5	886	12	CNS07JUX	AL614235 T3 end of
27	75.6	6.5	734	12	CNS00EVL	AL099163 Drosophila
28	75.2	6.4	1101	12	CNS00EVL	AL069526 Drosophila
29	75	6.4	1085	12	CNS016YR	AL107373 Drosophila
30	75	6.4	1190	12	CNS020N7	AL026508 Tetradon
31	74.8	6.4	876	12	CNS009G1	AL053529 Drosophila
32	74.8	6.4	1203	12	CNS015WU	AL106008 Drosophila
33	74.6	6.4	307	12	CNS00A3W	AL054893 Drosophila
34	74.6	6.3	1027	12	CNS02F50	AL212733 Tetradon
35	74.4	6.3	1101	12	CNS00EVL	AL071206 Drosophila
36	74.2	6.3	798	9	AL1909641	AL1909641 PM-BR217-
37	74	6.3	928	12	CNS00EVL	AL071063 Drosophila
38	73.8	6.3	996	12	CNS00EVL	AL071865 Drosophila
39	73.6	6.3	1101	12	CNS0042W	AL055440 Drosophila
40	73.6	6.3	1101	12	CNS00LT2	AL078714 Drosophila
41	73.6	6.3	905	12	CNS00KHX	AL077798 Drosophila
42	73.4	6.2	987	12	CNS014PQ	AL104456 Drosophila
43	73.2	6.2	1101	12	CNS003BD	AL064091 Drosophila
44	73	6.2	1001	12	CNS0155H	AL105023 Drosophila
45	72.8	6.2				

## ALIGNMENTS

RESULT 1  
LOCUS CNS0039G/c 1101 bp DNA linear GSS 03-JUN-1999  
DEFINITION Drosophila melanogaster genome survey sequence TET3 end of BAC #  
BACR08K10 of RPCI-98 library from Drosophila melanogaster (fruit  
fly), genomic survey sequence.

ACCESSION AL063921  
VERSION AL063921.1 GI:4941778  
KEYWORDS  
SOURCE  
ORGANISM

Drosophila melanogaster  
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;  
Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.

REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :  
BP 191 91006 Evry cedex - FRANCE (E-mail : seqref@genoscope.cns.fr  
- Web : www.genoscope.cns.fr)

COMMENT  
Determination of this BAC-end sequence was carried out as part of a  
collaboration with the Berkeley Drosophila Genome Project (BDGP).  
The BDGP is constructing a physical map of the Drosophila  
melanogaster genome using these BACs. For further information  
please see <http://www.fruitfly.org> The BDGP Drosophila  
melanogaster BAC library was prepared by Kazutoyo Osoegawa and  
Aaron Mammeter in Pieter de Jong's laboratory in the Department of  
Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,  
NY. The library is named RPCI-98 and was constructed by partial  
EcoRI digestion of Drosophila DNA provided by the BDGP from the  
isogenic strain y2; cn bw sp, the same strain used for the BDGP's  
P1 and EST libraries. A more detailed description of the library  
and how to order individual BAC clones, the entire library, or  
filters for hybridization from the BACPAC Resource Center can be  
found at [http://bacpac.med.buffalo.edu/drosophila\\_bac.htm](http://bacpac.med.buffalo.edu/drosophila_bac.htm).

FEATURES  
Source

1. 1101  
/organism="Drosophila melanogaster"  
/db\_xref="taxon:7227"  
/clone\_lib="RPCI-98"

BASE COUNT	201 a	64 c	131 g	202 t	503 others
ORIGIN	/clone="BACR08K10" /note="end : TET3"				

Query Match	8.48;	Score 99;	DB 12;	Length 1101;
Best Local Similarity	20.28;	Pred. NO. 4.7e-08;		
Matches 143;	Conservative 301;	Mismatches 261;	Indels 4;	Gaps 2;

Oy	2	gtrggattggtggaattgggatacctcccttaaccccttaaggaagcgtggaacaaag	61
Db	1097	RMGDDTWDRIOTKKDMDWMTKKMTMKDRADDRRMAGDDRRMMDGAGTWTATYMMWWW	1038
Oy	62	caagaagaagttaglcaactccctcccttataatataatataatgatcatgagagaca	121
Db	1037	WMTATDWTMDKMMWMTATKAQTDATWMTATRARMDWAGDRGAKRRDRRAATDADGAGR	978
Oy	122	tggtcatgataaggttlaataagagtagtltgattgataatgtccagcactagtttt	181
Db	977	DGRRRKKKKKDDDDKKGGKKKKAAKAAWATKMWMDWMDWMDKMDKADKDRADD	918
Oy	182	tgttctgtgattcttcataatgagcagcaaaatttataatataatataatgaataat	241
Db	917	DCAGSKDDDGKKKADDDDTGTRKDDDDKDKWMDMDKAKTGMWDATWMAATMMWGWAD	858
Oy	242	gattgattattctctcgttaatttltgtaaatagattaaacagccaaatgtgagtgaca	301
Db	857	ADWMTWDAADDDWMDRRDADWAMWMDMAWAGARTADRRMDGRRGKKGRKRRDRK	798
Oy	302	gttctcaaatgacacactgcactgagggcagtgtgatttccaatccaacccaattga	361
Db	797	ADKKDDAADDRDDAATWTTWTTTTR--DTDDMKKDTWMTWMAADRTWDRDDDDDRDA	740
Oy	362	aaactaaaataaaaagaattagattatataataattagttlaatccaaggttgcta	421
Db	739	GTGAKRRWRTWRRMKRRDTRMDDADADTDARDDRRRRODGDAGKGRKRRRRDR	680
Oy	422	atcaattatlaattaaacagatgatttltgataatttaattaaatttaatttga	481
Db	679	ATWDTDAMWADAMWTTTDTDTDDMDKDRRRKRGRRRRRTTARAAMWMTWKAMD-w	622
Oy	482	tttgaatgaactcaatlaatcacaaaaaacctlaatcaatlaataatcttaigtatata	541
Db	621	AKMDKTRDTRMDDRAADWTDRKADRRDMAKARBARBRABARADRRWTTGKTTT	562
Oy	542	atttgaataataaaattgataactttaaactctgaagttccctcataaaacacgctt	601
Db	561	ATWTTWMAAANAAMWAMWATTTATTTWTTWTTTWTWTTTWTWTTTAAWMAWMAWMTAT	502
Oy	602	aattggcctagattcaacgcctatcatccaactgcgcgagcaattataaataatga	661
Db	501	WAAWMTAAWMAAANAAMWATTTTTTTTTTTTTTAAWMTAAWMTWMTWTTWTTWMAAT	442
Oy	662	attatattcttttccaataaagcactccctaattgtttaaataatattgc	710
Db	441	TTTTTTTWTWMAATATTTTWTWTTTAAWMAAANAAMWMTWMTATATKCC	393

RESULT 2	
CNS00EVL	
LOCUS	1101 bp DNA linear GSS 04-JUN-1999
DEFINITION	Drosophila melanogaster genome survey sequence T7 end of BAC: BACR28923 of RPCT-98 library from Drosophila melanogaster (fruit fly), genomic survey sequence.
ACCESSION	AL069706
VERSION	AL069706.1
KEYWORDS	GI:4949849
SOURCE	GSS.
ORGANISM	fruit fly;
	Drosophila melanogaster
	Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
	Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
	Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.

REFERENCE 1 (bases 1 to 1101)  
AUTHORS Genoscope.  
TITLE Direct Submission  
JOURNAL Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :

**COMMENT**

sequestration of this BAC-end sequence was carried out as part of a collaboration with the Berkeley Drosophila Genome Project (BDGP). The BDGP is constructing a physical map of the *Drosophila melanogaster* genome using these BACs. For further information please see <http://www.fruitfly.org> The BDGP *Drosophila melanogaster* BAC library was prepared by Kazutoyo Oseegawa and Aaron Mamoser in Pieter de Jong's laboratory in the Department of Cancer Genetics at the Roswell Park Cancer Institute in Buffalo, NY. The library is named RPC1-98 and was constructed by partial EcoRI digestion of *Drosophila* DNA provided by the BDGP from the isogenic strain y2; cn bw sp, the same strain used for the BDGP's p1 and EST libraries. A more detailed description of the library and how to order individual BAC clones, the entire library, or filters for hybridization from the BACPAC Resource Center can be found at [http://bacpac.med.buffalo.edu/drosophila\\_bac.htm](http://bacpac.med.buffalo.edu/drosophila_bac.htm).

FEATURES	Location/Qualifiers
source	1. .1101

BASE COUNT	419 a	91 c	60 g	299 t	232 others
ORIGIN	/organism="Drosophila melanogaster" /db_xref="taxon:7227" /clone_lib="RPCT-98" /clone="BACR29B23" /note="end : 17"				

Query Match	8.0%	Score 94	DB 12	Length 1101
Best Local Similarity	35.7%	Pred. No. 3.6e-07		
Matches 195	Conservative 101	Mismatches 247	Indels 3	Gaps 1

OY	192	atctctctgtagagcgagcaaaattctatctatctatctatgaatgaatctgcttgat	251
Db	452	ATTTCCTCAHTTMMMMMMMAATMTWMMMAAATAATTATWATWMAAAWMMWMTT	511
OY	252	tctcgttaattctgtgaatagattcaaaacagctcaatgtggagtgaccagctgtcaat	311
Db	512	TWMMWMTATTTTWTTWANWTTATWTAATAAAATAATATTTAAWMAATATWTTAAWA	571
OY	312	gaccctcgactctgggcagtggtgtaatttccaatccacactcaattgaaactaaat	371
Db	572	TTTAAAMWMTATATTAATWTAATTAATMTWATTAATATAAAAAATTTTWTWATTAAT	631
OY	372	taaaaagattagattattaaattacttaggttaactcagggctggcctaactat	431
Db	632	TTTTTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA	691
OY	432	attaattaaacagatagatttttgataacttaattaaatttatttgatttgaatga	491
Db	692	AAAAAATAAAATAWMAWMAATWATWATTAATTAATAAAWMAATTAATAAAWMAATWAMA	751
OY	492	ctcaattcatcacaacaaacctaaccataactaactaactcgtgatacttaattgaat	551
Db	752	TAAATTTWATATATATWTTTWTWAMWMTTWAMWMTATATW---ATAWMAATTAATAWMAATAWMA	808
OY	552	ataaattgataactttaaacttcgaagttctctataaaaaacacgataaattggcta	611
Db	809	ATAMATAMATTAATAWMAWMAATWMTWMTATATWMAATATWMAWMAAATAATWATATWMAATW	868
OY	612	gatttaacagctattatccaacactggccaagcaattataaaactaataattatatt	671
Db	869	ATAAATAAAATAATWMTWMTWTTTWTWMAWMTATTAATAWMAWMTATAAAWMAAATAAAWMAA	928
OY	672	tttccaataaagactctcctaactgtttaaataatgtcctaaacctaataataaatt	731
Db	929	AATTAAMWMTWATATTTTATTAATAWMTATWMAATWMTWMTWMTATWMTATWMTATW	988
OY	732	attgat	737

[illegible]

Oy	474	ttttaggttggaagcaactcaattacccaaacaaaccttaaatatcatctatc	533
Db	800	WTTWTATTTTATATATATAWMTTWMTATMWMTAAWAAMATATATTATWMTATTTATTTT	741
Oy	534	gtgataaattcgaaaataaalgatataccctttaaacctgcagtttcctctaanaa	593
Db	740	WTTTATTTATTTTAAATWMTATWMTATWMTTWTWMTTTTTTTTTTTTTTTTTTTT	681
Oy	594	aacagtataatttggcgctagattcaacgcttatctcaaacctgcccagacaattatca	653
Db	680	TTATTTAATTTMMMAATWMTAATTTATTTAATTAATTTAAATTTATTTAAAAATTTVTMMA	621
Oy	654	aataaataattatatttttctctaagaacactccctcaattgttaaaaaataatgctcaa	713
Db	620	AAAATTTTATTTATTTATTAATWMTATTTAATTTATATATWMTTFAAATTTTAAATTA	561
Oy	714	aacactaataaanaattatcttgtagtctcttgccagtaggtagagaggtagctgacaata	773
Db	560	TTWMTTAAATTTATTTTATTTTAAATWMTAAWMTAAWMTAAWMTAAWMTAAWMTW	501
Oy	774	aattagtcacataaataataagatggtgctgctgtaaa	814
Db	500	TTTTTATTTATTAATTTTATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTT	460
RESULT	4		
CNSDNEO7			
LOCUS	CNS000607	1101 bp	DNA linear GSS 04-JUN-1999
DEFINITION	Drosophila melanogaster genome survey sequence Tetr3 end of BAC: BACR23P01 of RPCI-98 library from Drosophila melanogaster (fruit fly), genomic survey sequence.		
ACCESSION	AJ069440		
VERSION	AJ069440.1	GI:4943583	
KEYWORDS	GSS.		
SOURCE	fruit fly.		
ORGANISM	Drosophila melanogaster Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila. 1 (bases 1 to 1101)		
REFERENCE	Genoscope.		
AUTHORS	Direct Submission		
TITLE	Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :		
JOURNAL	Bp 191 91006 EVERY cexed - FRANCE (E-mail : seqref@genoscope.cns.fr		
COMMENT	- Web : www.genoscope.cns.fr  Determination of this BAC-end sequence was carried out as part of a collaboration with the Berkeley Drosophila Genome project (BDGP). The BDGP is constructing a physical map of the Drosophila melanogaster genome using these BACs. For further information please see http://www.fruitfly.org/The BDGP Drosophila melanogaster BAC library was prepared by Kazutoyo Oseegawa and Aaron Mammosser in Pieter de Jong's laboratory in the Department of Cancer Genetics at the Roswell Park Cancer Institute in Buffalo, NY. The library is named RPCI-98 and was constructed by partial EcoRI digestion of Drosophila DNA provided by the BDGP from the isogenic strain YZ; cn bw sp, the same strain used for the BDGP's pl and EST libraries. A more detailed description of the library and how to order individual BAC clones, the entire library, or filters for hybridization from the BACPAC Resource Center can be found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.		
FEATURES	Location/Qualifiers		
source	1..1101 /organism="Drosophila melanogaster" /db_xref="taxon:7227" /clone_lib="RPCI-98" /clone="BACR23P01" /note="end : Tetr3"		
BASE COUNT	366 a 66 c 104 g 351 t 214 others		
ORIGIN			
Query Match	7.8%, Score 91.4; DB 12; Length 1101;		









BASE COUNT	395 a	120 c	103 g	334 t	149 others
ORIGIN	/clone_11d="RPCT-98" /clone="BACR08K08" /note="end : TET3"				

Query Match	7.0%	Score 82.2	DB 12	Length 1101
Best Local Similarity	39.28%	Pred. No. 4.4e-05		
Matches 157; Conservative	57	Mismatches 186	Indels 0	Gaps 0

Oy 343 aaacccaacatcaattlgaaaactaaaaataaaagattgatatctaattattagg 402  
| : | | | | | | | | | | | | | | : | | |  
Db 675 AWAATTAAAAAATAAAAAAAAAATTTAAAWTWTAAATATATW 734

Oy 403 ttattcacggttcgctaaccatctattatctaataaacgactgattcttgtaaatc 462

Db 735 TTTTAAATAATATATTAAAAAAWATATTTTATTATATTAAAAATATATATTTTATA 794

```
Oy 463 taattaaaatttatttgattgaaactcaattacacaaaaaacctaat 522
      || : : || : || : : : : || : : || : || : ||
Db 795 TATNNNNATATATTTTAAATTWATATWTTAAATWTTAAAWATTNTAAATAAAA 854
```

Oy 523 taatatctatgtgataaattagaaatataaatgatctaaccttaacctcgagttc 582  
::| | ||||| : : ::| : | || :  
Db 855 WAAAAAAAAATAAATTAATTATTTWATAMWT TWWAAATAAATTAATTAATAAAAAA 914

Qy 583 tctataaaaaaacacgtataattggtactagattaacgcattattcaaacctgcagg 642  
||| : ||| | : : ||| : : : |||  
Db 915 AATTAAWATTAATAATWATTTWTTMTYTAAAAWATWATTAAWAWAAATATTTTAT 974

**Oy** 643 acaaattatcaaaaataataattatttttllctaataagcacttcctaatgttaaaa 702  
| :||:|||| ||| :: || | :| : | : : ||||  
**Db** 975 ATWATWATATTAATAWWAAAAAAATTAATAAWAAAATTNWAAMAWAAYAAAWAYACWAAAA 103

OY	703	tatatgtcctaaacactataataaaattatgtgtaac	742
		::   ::   ::   ::   ::   :	
Db	1035	AATMTATATATATAAAAAAAHTAANTTATTATTAAGAMANN	1074

```
RESULT_10  
CNSOLG8P  
  
.....
```

**DEFINITION** Anopheles gambiae GSS sp6 end of clone 06E14 of Notredame1 library from strain PEST of Anopheles gambiae (African malaria mosquito), genomic survey sequence.

VERSION	GI:7000944
KEYWORDS	ALL42826.1
SOURCE	GSS
ORGANISM	African malaria mosquito.

Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta; pterygota; Neoptera; Endopterygota; Diptera; Nematocera; Culicoidae; Anopheles.

Genoscope - Centre National de Séquençage :  
 Submitted (16-FEB-2000)  
 Direct Submission  
 Genoscope.  
 AUTHORS  
 JOURNAL

**REFERENCE**  
-Web : [www.genoscope.cns.fr](http://www.genoscope.cns.fr)  
2 (bases 1 to 910)  
**AUTHORS**  
Roth,C.W., Brey,P.T., Ke,Z., Collins,F.H. and Weissenbach,J.  
**TIME**  
Direct Submission

COMMENT  
JOURNAL  
Submitted (16-FEB-2000) BBMI, Institut Pasteur, 25, rue du Dr.  
Roux, Paris 75015, France  
This clone is from an *A. gambiae* BAC library provided by F. H.  
Collins and sequenced by Genoscope. It is collinear with the  
BAC clone.

FEATURES	source
Location/Qualifiers	1 010
Laboratory Of Biochem. and Biol. Molec. Of Insects, Institut Pasteur.	

```

/organism="Anopheles gambiae"
/strain="PEST"

```

```

      .
      /db_xref="taxon:7165"
      /clone="06E14"
      /clone_id="Notredame1"
      /note="end : Sp6"
BASE COUNT      376 a      83 c      91 g      294 t      66 others
ORIGIN

```

Query Match	6.9%	Score 81.2	DB 12	Length 910
Best Local Similarity	42.8%	Pred. No. 6.8e-05		
Matches 188	Conservative 30	Mismatches 221	Indels 0	Gaps 0

**Qy**    349 caactcaattgaaacataaaaatgaattagatatataatctagtgttaatt 408  
        | : | | | | : | | | | | : || |:| :| :| | | :|  
**Db**    472 CWTGTTTAAATTAATAATTNNAAAAAAMWMTTWWATZATAAMTTTTWAAA 531

QY 409 caccggttcgctaaccattatctaataaaccgatactlltggatcaatttaatta 468

Db 532 AAATATATATATTTTTPAANAAATWTTAAATATATATWTTAAATTTAAANAAW 591

Qy 469 aaatttattggaattggaatgaactcaatlacatcacaaaaaacctaataatcat 528  
:  
|| || | : : : ||||| || ||  
Db 592 WAAATTTTATTATTAWTAA TTWANA AAAAATAATAAAAAAAAANNTAT 651

QY 529 cctatgtgataataatttagaataataatgatctaaccctttaaactctcagatttccttat 5808  
:: | ||| : ||| | ||| : | ||| : | | |  
Db 652 wwwwtttaattatataaawaaaaattttgataatwtttptfaaaaaawawtttwaattwact 7111

**Qy** 589 aaacaacagctataatgtggcgcagattcaacgctatttcacactgcccagacaatt 648  
| : | | | | | : | | | | |  
**Db** 712 TATWATATAAAAAATATTWAAAAAATAATTAATATAAAATAANAAAAAAAAATA 771

Qy 649 attcaaatcataatattatcttttccataagaagcacttcctaattgttcaaatatctg 708  
 | : ||| | | : | | | | | : | | :  
 Db 772 NAAATTNWTAAAAAATTWANNANAATTAYTWACAAWNTNTWTATAAAAATATNTYTAA 831

**DY**      tcctaacaccataataaaattatctgtgtaaccttggcagtcggtgagaggtgcgcac 768  
         ||| : ||| : | | | | | | | | | |  
**Db**      AATAAWTTTAAATWTNNATTATTCATTAATTTTAAATAATTATATTAATAAATAATTAAT 891

```
QY 769 aaataatgtagcaca 787
      | | | | | : | | | |
Db 892 ATCTAATATATATATAA 910
```

RESULT 11  
CNS010MP/c

**DEFINITION**  
Drosophila melanogaster genome survey sequence 17 end of BAC BACN04L20 of DrosBAC library from *Drosophila melanogaster* (fruit fly) genomic survey sequence.

ORCAUTREX  
SOURCE  
KEYWORDS  
VERSION  
AL099163.1  
fruit fly.  
GSS.  
GI:5610774  
repressor  
breakthru

Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta; pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila. 1 (base) 1 + 5 (734)

**GENOSCOPE.**  
**AUTHORS**  
**TITLE**  
**Direct Submission**  
**Submitted (23-JUL-1999)**  
**Genoscope - Centre National de Séquençage :**  
**BP 101 91106 Evry cedex - FRANCE / E-mail : [contact@genoscope.ces.fr](mailto:contact@genoscope.ces.fr)**

COMMENT  
-web : [www.genoscope.cns.it](http://www.genoscope.cns.it)  
determination of this BAC-end sequence was carried out as part of a  
collaboration with the European Drosophila Genome Project (EDGP) -  
<http://www.edgp.chi.ac.uk> - This Drosophila melanogaster BAC

project grant. The DNA was prepared from embryos by Alain Bucheton and Genevieve Payan. It has been constructed in the vector pBel0BAC11.



[illegible]

QY	726	aaattatttgatcttggcagta	752
Db	1008	TATATTATAGMAATATATAAAATA	1034
<hr/>			
RESULT 14			
CNS04DOK LOCUS DEFINITION	CNS04DOK Tetraodon nigroviridis genome survey sequence T7 end of clone 101H21 of library G from Tetraodon nigroviridis, genomic survey sequence.	945 bp	DNA linear GSS 21-MAY-2000
ACCESSION VERSION KEYWORDS SOURCE ORGANISM	AL285149 GI:8023560 GSS: genome survey sequence. Tetraodon nigroviridis. Tetraodon nigroviridis Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Acanthopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei; Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes; Tetraodontidae; Tetraodon.		
REFERENCE AUTHORS	1 (bases 1 to 945) Roest-Crollius,H., Jalllon,O., Dasilva,C., Fitzames,C., Fisher,C., Bouneau,L., Billault,A., Quetier,F., Saurin,W., Bernot,A. and Weissenbach,J.		
TITLE	Characterization and repeat analysis of the compact genome of the freshwater pufferfish Tetraodon nigroviridis		
JOURNAL REFERENCE AUTHORS	Unpublished 2 (bases 1 to 945) Roest-Crollius,H., Jalllon,O., Dasilva,C., Bouneau,L., Fisher,C., Bernot,A., Fitzames,C., Wincker,P., Broitler,P., Quetier,F., Saurin,W. and Weissenbach,J.		
TITLE	Human gene number estimate provided by genome wide analysis using Tetraodon nigroviridis DNA sequence		
JOURNAL REFERENCE AUTHORS TITLE JOURNAL COMMENT	Unpublished 3 (bases 1 to 945) Genoscope. Direct Submission Submitted (12-Apr-2000) to the EMBL/Genbank/DBJ databases This sequence is a single read and was generated as part of a large scale clone-and sequencing project of the Tetraodon nigroviridis genome. For more information, please take a look at <a href="http://www.genoscope.cns.fr/Tetraodon">http://www.genoscope.cns.fr/Tetraodon</a> .		
FEATURES	Location/Qualifiers		
source	1..945 /organism="Tetraodon nigroviridis" /db_xref="taxon:99883" /clone="101H21" /clone_1lb="G" /note="Genoscope sequence ID : COBG10ICD1LP1-end : T7"		
BASE COUNT	386 a     112 c     96 g     231 t     120 others		
ORIGIN			
<hr/>			
Query Match	6.7%; Score 78.4; DB 12; Length 945;		
Best Local Similarity	41.5%; Pred. No. 0.00021;		
Matches	192; Conservative 49; Mismatches 215; Indels 7; Gaps 1;		
QY	336	attttcaaatcacactcaatttgaacactaaaattaacagaattagattttaaat	395
Db	490	ATTTTWAAAAAMAAAATATAAAAAATAWMMWTTTTTAAAAAAAAMATATTAAMATTWTGA	549
QY	396	tatttggttaattcaacgggttgccatacatattatatatttaattaaacgatagtatttt	455
Db	550	MAMMATTTTTWATATTTTTTTATTTAAAMAATATATATATAAAAAATTTTAAAAAAAAMATATTTTTT	609
QY	456	gataatttaataaatttataggattgaaatgaactcaatcaaacacaacaaacta	515
Db	610	AAMAAATTAAMATATAAAATATAATATATTATATAAAMAMATAMAMMMMTTAMMAAMMMTMA	669
QY	516	atcaaatataatcttatgtgatataaatttgaaatatataagattcaaccttaaatctc	575
Db	670	ATATATAAAATATAAAMATATAAAATATAAAATATAAAMAMMMMTTATATAAANAANA	729

QY	576	gagtttcctcttaaaaaaacgcyataatttggcgcagatttcaagccatttcaact	635
Db	730	AAAAAAAAATATATATATWAAA-----ATATATATTTAAATATATAATATWATATAATAA	782
QY	636	ggccgacacatttataaataatcaatatttcttcttaataagcaacttccat	695
Db	783	AAAAAAAAATTTTATATATATAAAATTAAMWATATATATTAANATTAANNTTAAMNTAM	842
QY	696	gttaaatatattgtcttaacactataataaatttatttgcgtatcttggcagtagt	755
Db	843	NTAATTAATTAATTAATTTAAAAAATAAATAAATTTATATAAATATTAATTAATAATAA	902
QY	756	ggaggtgcctgcacaaataaattagtcataaataatgat	798
Db	903	TATATAAAAAAAAAAAAAAAAAATTTTAAAAAATAAATAAATAATAT	945
RESULT 15			
CNS0039L/c			
LOCUS			
DEFINITION	CNS0039L	1101 bp	DNA linear GSS 03-JUN-1999
ACCESSION	Drosophila melanogaster genome survey sequence T7 end of BAC #		
VERSION	BACR0810 of RPCT-98 library from Drosophila melanogaster (fruit		
KEYWORDS	fly), genomic survey sequence.		
SOURCE	AL063926		
ORGANISM	AL063926.1	GI:4941783	
	GSS.		
	fruit fly.		
	Drosophila melanogaster		
	Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;		
	Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;		
	Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.		
REFERENCE	1 (bases 1 to 1101)		
AUTHORS	Genoscope.		
TITLE	Direct Submission		
JOURNAL	Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :		
	BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr		
	- Web : www.genoscope.cns.fr)		
COMMENT	Determination of this BAC-end sequence was carried out as part of a		
	collaboration with the Berkeley Drosophila Genome Project (BDGP).		
	The BDGP is constructing a physical map of the Drosophila		
	melanogaster genome using these BACs. For further information		
	please see <a href="http://www.fruitfly.org">http://www.fruitfly.org</a> The BDGP Drosophila		
	melanogaster BAC library was prepared by Kazuhiro Osoegawa and		
	Aaron Määmøser in Pieter de Jong's laboratory in the Department of		
	Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,		
	NY. The library is named RPCT-98 and was constructed by partial		
	EcoRI digestion of Drosophila DNA provided by the BDGP from the		
	isogenic strain Y2; cn bw sp, the same strain used for the BDGP's		
	p1 and EST libraries. A more detailed description of the library		
	and how to order individual BAC clones, the entire library, or		
	filters for hybridization from the BACPAC Resource Center can be		
	found at <a href="http://bacpac.med.buffalo.edu/drosophila_bac.htm">http://bacpac.med.buffalo.edu/drosophila_bac.htm</a> .		
FEATURES			
source			
	1..1101		
	Location/Qualifiers		
	/organism="Drosophila melanogaster"		
	/db_xref="taxon:7227"		
	/clone.lib="RPCT-98"		
	/clone="BACR08110"		
	/note="end : T7"		
BASE COUNT	139 a	79 c	186 g
			360 t
			337 others
ORIGIN			

Db	1007	RADRRADATARRKKTKTGMTGARATRTKTRTMTWTRBMAAGTRATTKMGDDTAPRTWCT	948
OY	463	taattaaatttcttgygatttgaatgaatgaactcaatcaatcaacacaactcaatcaat	522
Db	947	TTATATCTCKKCTATWCKGTGCTTKTKTKTKRGCTRKTKTKKAMMAAAAAAAMAMW	888
OY	523	taattctctatgatataatttgaataatgaataatgaatgaatgaatgaatgaatgaatc	582
Db	887	TTATATATDTVKMTATGMAWATTKADATADWMAWDTKTKTKTATATTTDKDAAAAAA	828
OY	583	tctataaaaaacacgataatgygctagatlttaacagctatalttcaactgygcag	642
Db	827	AARATGAAAAARAAAAARAKMTDDRWMAAAMDAAAARDARGGTGAARDGMCGKKAAK	768
OY	643	acaattatataatcaatattatlttcttcaataaagcacttccaatgtgtataa	702
Db	767	WTGRTTAAAAADMAAAAAAAATMAMMAAAATRAKTAAAAAATATMGMAAAATKTKAAAA	708
OY	703	tatattcttaacgctaataataatattatttgytacttgygcagtgygagaggt	762
Db	707	ADAWAARAGDKTAAAAAARARARARAGSKRTRKKAARAAATAAAAATAAAAAADAAD	648
OY	763	gctacaataatctagtcgata-----aaatataatgaattgtygtctgtganaag	815
Db	647	AMTKKKAMMARTTTTPTGCKTKCTKCTGSKGAAMWAGTATKKKAAAAAKTKMAAAAAA	588
OY	816	acaggtgaggaagaagccacctctcgaatgcataaagaagccalttcaacaacccaat	875
Db	587	AAAGKRTGKKKKKRTBTBTRDMNNAHAATMNAATBMGTGKSMNMAATMDMTWTKG	528
OY	876	gggagccacacacgctcccgccatlaaatccatccatctcaacaaacaaactccag	935
Db	527	KRGGRGGRKTKKYYTTMMVPTGCCCCCCKCAAKKTTTSGCBTBTYCMMAAAACNCCMCA	468
OY	936	attcttccacaagaacagatlttccaatgaatglttccctataactaaccccccaac	995
Db	467	MTMAMMCCAMMMMMCMCMAMACANNTMCAAMCMCMACACACMCCMMAACCMACMACA	408
OY	996	aactcataataaccgaatttgccttccaccaaaccgcgctccgctgcgcagccaattct	1055
Db	407	CMCCAACCCACATCMCMCMCCMMHTTMAACCAATATMAAAAAACCMHTTMAACTCAAMCC	348
OY	1056	ataatcagcaggaat	1069
Db	347	CTCTMAAMMAACT	334

Search completed: July 30, 2002, 10:38:02  
Job time: 5471 sec

	Query Match	Similarity	7.1%	Score	78.4	DB	12	Length	1101
	Best Local	Similarity	27.1%	Pred.	No.	0.00021			
	Matches	199	Conservative	185	Mismatches	343	Indels	7	Gaps
QY	343	aaatcacactcaatttggaacactaaataaagaatttgattatataattattagg	402						
DB	1067	AAKRRPAAAFRIAMAMATTAATAAKKKWRRAAMPAAMWDGATDITATMDGDGDDRRDAADR	1008						
QY	403	ttaattcagaggttgcgtcaatcaattatattataaagaatagatlatatttggaatt	462						

---





GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 30, 2002, 10:04:36 ; Search time 50.16 Seconds  
(without alignments)  
5739.284 Million cell updates/sec

Title: US-09-530-663b-5

Perfect score: 1172

Sequence: 1 tgttagattgtgtggaatggg.....agcccgcaatgacgcgcgaca 1172

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 383533 seqs, 122816752 residues

Total number of hits satisfying chosen parameters: 767066

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Issued\_Patents\_NA:\*  
1: /cgn2\_6/ptodata/1/ina/5A\_COMB.seq:\*  
2: /cgn2\_6/ptodata/1/ina/5B\_COMB.seq:\*  
3: /cgn2\_6/ptodata/1/ina/6A\_COMB.seq:\*  
4: /cgn2\_6/ptodata/1/ina/6B\_COMB.seq:\*  
5: /cgn2\_6/ptodata/1/ina/PTC05\_COMB.seq:\*  
6: /cgn2\_6/ptodata/1/ina/backfiles1.seq:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	61.8	5.3	636	4	US-08-998-416-1137 Sequence 1137, App
2	61.8	5.3	665	2	US-08-883-795A-36 Sequence 36, Appl
3	61.4	5.2	660	1	US-07-991-867B-32 Sequence 32, Appl
4	61.4	5.2	660	1	US-08-107-755A-32 Sequence 32, Appl
5	61.4	5.2	660	2	US-08-544-332-32 Sequence 32, Appl
6	61.4	5.2	1511	1	US-07-991-867B-8 Sequence 8, Appl
7	61.4	5.2	1511	1	US-08-107-755A-8 Sequence 8, Appl
8	61.4	5.2	1511	2	US-08-544-332-8 Sequence 8, Appl
9	61.4	5.2	4810	3	US-08-852-629-11 Sequence 11, Appl
10	61.4	5.2	4838	3	US-08-852-629-15 Sequence 15, Appl
11	61.2	5.2	837	4	US-08-998-416-288 Sequence 288, App
12	60	5.1	1431	4	US-09-316-083-2 Sequence 2, Appl
13	60	5.1	6243	2	US-09-056-075-1 Sequence 1, Appl
14	58.8	5.0	636	4	US-08-998-416-1137 Sequence 1137, App
15	58.8	5.0	837	4	US-08-998-416-288 Sequence 288, App
16	58.2	5.0	615	4	US-08-998-416-186 Sequence 186, App
17	57.6	4.9	2317	4	US-08-749-522-5 Sequence 5, Appl
18	57.6	4.9	3974	2	US-08-467-504-3 Sequence 3, Appl
19	57.2	4.9	665	4	US-08-883-795A-36 Sequence 36, Appl
20	57.2	4.9	3095	6	5231168-1 Patent No. 5231168
21	56.8	4.8	615	4	US-08-998-416-186 Sequence 186, App
22	56.6	4.8	2317	3	US-08-749-522-5 Sequence 5, Appl
23	55.6	4.7	3974	4	US-08-467-504-3 Sequence 3, Appl
24	55.6	4.7	724	4	US-08-998-416-683 Sequence 683, App
25	55.4	4.7	854	4	US-08-998-416-534 Sequence 534, App
26	55.4	4.7	860	4	US-08-998-416-287 Sequence 287, App

C 28	54.6	4.7	663	4	US-08-998-416-191	Sequence 191, App
C 29	54.6	4.7	1186	2	US-08-731-722-5	Sequence 5, Appl
C 30	54	4.6	19124	2	US-08-487-826B-13	Sequence 13, Appl
C 31	53.6	4.6	854	4	US-08-998-416-534	Sequence 534, App
C 32	53.6	4.6	19124	2	US-08-487-826B-13	Sequence 13, Appl
C 33	53.4	4.6	3926	2	US-08-731-722-1	Sequence 1, Appl
C 34	53.4	4.6	3926	2	US-08-731-722-1	Sequence 1, Appl
C 35	53.4	4.6	3926	2	US-08-731-722-2	Sequence 2, Appl
C 36	53.4	4.6	3926	2	US-08-731-722-2	Sequence 2, Appl
C 37	53	4.5	658	4	US-08-998-416-595	Sequence 595, App
C 38	52.8	4.5	658	4	US-08-998-416-595	Sequence 595, App
C 39	52.4	4.5	2251	4	US-08-991-677-11	Sequence 11, Appl
C 40	52.4	4.5	6124	4	US-08-213-419B-3	Sequence 3, Appl
C 41	52.2	4.5	701	4	US-08-998-416-701	Sequence 701, App
C 42	52.2	4.5	1422	1	US-08-319-704-5	Sequence 5, Appl
C 43	52	4.4	724	4	US-08-998-416-683	Sequence 683, App
C 44	52	4.4	732	4	US-08-998-416-1036	Sequence 1036, App
C 45	52	4.4	732	4	US-08-998-416-1036	Sequence 1036, App

#### ALIGNMENTS

RESULT 1  
US-08-998-416-1137  
Sequence 1137, Application US/08998416  
Patent No. 6239264  
GENERAL INFORMATION:  
APPLICANT: Philippesen, Peter  
APPLICANT: Pohlmann, Rainer  
APPLICANT: Steiner, Sabine  
APPLICANT: Mohr, Christine  
APPLICANT: Wendland, Jurgen  
APPLICANT: Knechtle, Philipp  
APPLICANT: Reibschung, Corinne  
TITLE OF INVENTION: GENOMIC DNA SEQUENCES OF ASHBYA GOSSPYII  
TITLE OF INVENTION: AND USES THEREOF  
NUMBER OF SEQUENCES: 1152  
CORRESPONDENCE ADDRESSES:  
ADDRESSEE: No. 6239264artis Corporation  
STREET: 3054 Cornwallis Road  
CITY: Research Triangle Park  
STATE: No. 6239264th Carolina  
COUNTRY: USA  
ZIP: 27709  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/998,416  
FILING DATE: 24-DEC-1997  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: CH 0016/97  
FILING DATE: 31-DEC-1996  
ATTORNEY/AGENT INFORMATION:  
NAME: Meigs, J. Timothy  
REGISTRATION NUMBER: 38,241  
REFERENCE/DOCKET NUMBER: PE/5-30306/A/CCG1976  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 919-541-8587  
TELEFAX: 919-541-8689  
INFORMATION FOR SEQ ID NO: 1137:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 636 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
ORIGINAL SOURCE:  
ORGANISM: PAG1692ARP







```
; LENGTH: 1511 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: unknown
; MOLECULE TYPE: DNA (genomic)
; ORIGINL SOURCE:
; ORGANISM: Amsacta moorei entemopoxvirus
; FEATURE:
; NAME/KEY: CDS
; LOCATION: complement (18..218)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: complement (234..782)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 852..1511
US-07-991-867B-8
```

```
Query Match 5.2%; Score 61.4; DB 1; Length 1511;
Best Local Similarity 48.0%; Pred. No. 5e-05;
Matches 205; Conservative 0; Mismatches 221; Indels 1; Gaps 1;
```

```
QY 369 aattaaaaagatttagatttaataatttagttaattcagggctgctcaatcaalt 428
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 942 AATATTATATATTAATAAAATTTAGTTAATTAGAAAGATTGCATATATATATGAT 1001

QY 429 attattaataaagcagtagtattttagataatttaataaaatttagattgaat 488
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1002 AATTAATATTTTAATAATATTCAGAAAATATTAAGTTTATATATTTCAATTTAAAT 1061

QY 489 gaaccaattacatcacaaaaaccatcaatcaataatactcattg-ataaatttag 547
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1062 ATTATTATTTAATTTATATACAAATTAATAAAATATACATATTTAGATATATCTTAT 1121

QY 548 aaataataatgattaaaccttaaatctcgagttctctataaaaaacgcataattgg 607
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1122 AACCAAAATAGCAATATAGTAATTTATACACCATTTCTATAGAAATTTTAAATTTGT 1181

QY 608 gctagatttaacagctattatcaaacctgcagagacaattttaataataattatt 667
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1182 GAATCATGTGAATATATACATATATTTTATTAATTAATTTTAAATTTAAATAATTA 1241

QY 668 atttttcttaataaagcacttccttaattglttaaatatatactgaacactaataaa 727
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1242 ATTAATATCTAAATAATTAATTTGTAACCTTATATATGTTTCCATATGTATAGTTGAG 1301

QY 728 attatttgctatcttggcagtagtgagagtgctgacaataatttagtgcataaa 787
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1302 TTAAATATGGAATCAATACAAATAAAGATTTAATTTATGAAAAAATTAAATTAATTTA 1361

QY 788 atataat 794
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1362 AAAAAAT 1368
```

```
RESULT 7
US-08-107-755A-8
; Sequence 8, Application US/08107755A
; Patent No. 5721352
; GENERAL INFORMATION:
; APPLICANT: Moyer, Richard W.
; APPLICANT: Hall, Richard L.
; APPLICANT: Gruidl, Michael E.
; TITLE OF INVENTION: No. 5721352el Entomopoxvirus Expression System
; NUMBER OF SEQUENCES: 40
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: David R. Saliwanichik
; STREET: 2421 N.W. 41st Street, Suite A-1
; CITY: Gainesville
; STATE: Florida
; COUNTRY: U.S.A.
; ZIP: 32606
```

```
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/107,755A
; FILING DATE: 19-AUG-1993
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/827,658
; FILING DATE: 30-JAN-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/657,584
; FILING DATE: 19-FEB-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: Saliwanichik, David R.
; REGISTRATION NUMBER: 31,794
; REFERENCE/DOCKET NUMBER: UF114.C2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (904) 375-8100
; TELEFAX: (904) 372-5800
; INFORMATION FOR SEQ ID NO: 8:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1511 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: unknown
; MOLECULE TYPE: DNA (genomic)
; ORIGINAL SOURCE:
; ORGANISM: Amsacta moorei entemopoxvirus
; FEATURE:
; NAME/KEY: CDS
; LOCATION: complement (18..218)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: complement (234..782)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 852..1511
US-08-107-755A-8
```

```
Query Match 5.2%; Score 61.4; DB 1; Length 1511;
Best Local Similarity 48.0%; Pred. No. 5e-05;
Matches 205; Conservative 0; Mismatches 221; Indels 1; Gaps 1;
```

```
QY 369 aattaaaaagatttagatttaataatttagttaattcagggctgctcaatcaalt 428
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 942 AATATTATATATTTAATAAAATTTAGTTAATTAGAAAGATTGCATATATATATGAT 1001

QY 429 attattaataaagcagtagtattttagataatttaataaaatttagattgaat 488
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1002 AATTAATATTTTAATAATATTCAGAAAATATTAAGTTTATATATTTCAATTTAAAT 1061

QY 489 gaaccaattacatcacaaaaaccatcaatcaataatactcattg-ataaatttag 547
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1062 ATTATTATTTAATTTTATTAACAAATTTAAATAATTAACATATTTAGATATATCTTAT 1121

QY 548 aaataataatgattaaaccttaaatctcgagttctctataaaaaacgcataattgg 607
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1122 AACCAAAATAGCAATATAGTAATTTATACACCATTTCTATAGAAATTTTAAATTTGT 1181

QY 608 gctagatttaacagcactataatcaaacctgcagagacaattttaataataattatt 667
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1182 GAATCATGTGAATATATACATATATTTTATTAATTAATTTTAAATTTAAATAATTA 1241

QY 668 atttttcttaataaagcacttccttaattglttaaatatatactgaacactaataaa 727
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1242 ATTAATATCTAAATAATTAATTTGTAACCTTATATATGTTTCCATATGTATAGTTGAG 1301

QY 728 attatttgctatcttggcagtagtgagagtgctgacaataatttagtgcataaa 787
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
```









	Query Match	5.1%: Score 60; DB 2; Length 6243;	
	Best Local Similarity	45.8%: Pred. NO. 0.00013;	
	Matches 207; Conservative	0; Mismatches 245; Indels	Gaps 0;
QY	178	ttttgttcgtgctgattcttcgatgagcgaaatttcatatatataatgaata	237
Db	2819	TTTTTANGTTTAAATTATATATTTTATAAATTTTAAACATAATTAATGATTCT	2878
QY	238	ataatgatgatattctctgcgaatttgcgaatagatnaaacagctcaatgtagtg	297
Db	2879	TTTTTAATTTTTTACATTTTCATTCATTAATATATATACATATATTTTCAATATATAT	2938
QY	298	accagctgcgaatgacacactgcactgcgtggggcagtgagatttccaatcacacactcaat	357
Db	2939	TCTTCATTTGTAAATTTTATAGATGATTTTACTATTTTACGTTTATATATATATTAATATTA	2998
QY	358	ttgaaactaaatttaaaaagatttagatataatataattaggttaattcaagggctg	417
Db	2999	TGTAAATTTATATATAAATTCMAAGGAGCTTATTAATATATATATTCCAAAGTACT	3058
QY	418	gctaatcaattatataatnaaacgagtagatttgataatataatnaaatattcat	477
Db	3059	AAAGATTAATTTTTTTCATTTTAAACAATACTTTTGTATATATATGTTTAATTAATTAAT	3118

Db 3239 GTGAATTTCCGGCTTATTAACTTATTTT 3270

US-08-998-416-1137

Db 498 TTTAATAATAGAAATTAAAGTTAAATTTTAAATAATAATTC TTATAAAAAGATTAA 4399

```

QY 397 attaggttaattcaggggttggtcctaatacattatattatataaagaagatgattttg 456
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 438 ATAAATATAATCAACATATATTTATATAAAATAGATATTAATAATATAATATATTAC 379
QY 457 ataatttaataaatttatttgatttggaatgaactcaattacacaaaccctaa 516
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 378 AATATTTAAATTAATTAATCTTTAATTAATAATATATATATTTTAAATCAATATTA 319
QY 517 tcaaatataatcttaigtgataataattagaataataatgattacacttaaacctcg 576
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 318 TAAATTAATATTTAATTTTGTGATATCTATTTAATTAATTTTAAAGAAATTAATAAT 259
QY 577 agttctcttaataaaacacgtataaattggcgtagatttaacagctatttcaactg 636
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 258 CTAATTAATATTTTAACTAATTTAAATTTGAACATAGACTAAATAGTATTCATATTA 199
QY 637 gccaggacaattataaat-taataattatttttcttaataaagcacttctaatt 695
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 198 AATATTTAATTTAATTAATTAATTAATTAATGAATGAATTAATTAATTAATTAATTA 139
QY 696 gtaaaataatgtcttaaacactaataataaattatttgtaactttggcagtagt 755
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 138 ATTAATATAATAGTATTTAATTAATCAAAATTAATTTTATTAATTAATGAATTAATAGT 79
QY 756 gagagggtcgcacaataaattagtgacataaataatggatt 799
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 78 TAAATTAATTTAATCTTAATATATTTAATAAAGTTTATATTT 35

```

```

RESULT 15
US-08-998-416-288/c
; Sequence 288, Application US/08998416
; Patent No. 6239264
; GENERAL INFORMATION:
; APPLICANT: Philippen, Peter
; APPLICANT: Pohlmann, Rainer
; APPLICANT: Steiner, Sabine
; APPLICANT: Mohr, Christine
; APPLICANT: Wendland, Jurgen
; APPLICANT: Knechtle, Philipp
; APPLICANT: Reibschung, Corinne
; TITLE OF INVENTION: GENOMIC DNA SEQUENCES OF ASHBYA GOSYPYII
; TITLE OF INVENTION: AND USES THEREOF
; NUMBER OF SEQUENCES: 1152
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: No. 6239264artis Corporation
; STREET: 3054 Cornwallis Road
; CITY: Research Triangle Park
; STATE: No. 6239264th Carolina
; COUNTRY: USA
; ZIP: 27709
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentln Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/998,416
; FILING DATE: 24-DEC-1997
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: CH 0016/97
; FILING DATE: 31-DEC-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Meigs, J. Timothy
; REGISTRATION NUMBER: 38,241
; REFERENCE/DOCKET NUMBER: PE/5-30306/A/CGC1976
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 919-541-8587
; TELEFAX: 919-541-8689
; INFORMATION FOR SEQ ID NO: 288:
; SEQUENCE CHARACTERISTICS:

```

```

; LENGTH: 837 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; ORIGINAL SOURCE:
; ORGANISM: PAGI24LRP
; US-08-998-416-288

```

```

Query Match 5.0%; Score 58.8; DB 4: Length 837;
Best Local Similarity 48.3%; Pred. No. 0.00016;
Matches 224; Conservative 0; Mismatches 237; Indels 3; Gaps 2;

```

```

QY 339 ttcaaatcacacactcaattgaaacactaaatbaaaagatttagatttaaat-- 396
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 498 TTAATTAATAGAAATTAAGTTAAATTAATTTTATATATATATCTTTTAAATAAATTA 439
QY 397 attaggttaattcaggggttggtcctaatacattatattatataaagaagatgattttg 456
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 438 ATAAATATAATCAACATATATTTTAAATAATAGATATTAATAATTAATTAATTAATTA 379
QY 457 ataatttaataaatttatttgatttggaatgaactcaattacacaaaccctaa 516
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 378 AATATTTAATTAATTTAATCTTTAATTAATAATTAATTAATTTTATTAATAATTAATTA 319
QY 517 tcaaatataatcttaigtgataataattagaataataatgattacacttaaacctcg 576
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 318 TAAATTAATATTTAATTTTGTGATATCTATTTAATTAATTTTAAAGAAATTAATAAT 259
QY 577 agttctcttaataaaacacgtataaattggcgtagatttaacagctatttcaactg 636
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 258 CTAATTAATATTTTAACTAATTTAAATTTGAACATAGACTAAATAGTATTCATATTA 199
QY 637 gccaggacaattataaat-taataattatttttcttaataaagcacttctaatt 695
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 198 AATATTTAATTTAATTAATTAATTAATTAATGAATGAATTAATTAATTAATTAATTA 139
QY 696 gtaaaataatgtcttaaacactaataataaattatttgtaactttggcagtagt 755
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 138 ATTAATATAATAGTATTTAATTAATCAAAATTAATTTTATTAATTAATGAATTAATAGT 79
QY 756 gagagggtcgcacaataaattagtgacataaataatggatt 799
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 78 TAAATTAATTTAATCTTAATATATTTAATAAAGTTTATATTT 35

```

```

Search completed: July 30, 2002, 11:11:03
Job time: 3987 sec

```

---

**THIS PAGE BLANK (USPTO)**

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 29, 2002, 21:58:04 ; Search time 1921.77 Seconds  
(without alignments)  
174.227 Million cell updates/sec

Title: US-09-530-663B-15  
Perfect score: 16  
Sequence: 1 ccttcaccaaccccc 16

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 1797656 segs, 10463268293 residues  
Total number of hits satisfying chosen parameters: 3595312

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : GenBank1:  
1: gb\_ba:\*  
2: gb\_hgt:\*  
3: gb\_in:\*  
4: gb\_cm:\*  
5: gb\_ov:\*  
6: gb\_pat:\*  
7: gb\_ph:\*  
8: gb\_pl:\*  
9: gb\_pr:\*  
10: gb\_ro:\*  
11: gb\_sts:\*  
12: gb\_sy:\*  
13: gb\_un:\*  
14: gb\_vl:\*  
15: em\_ba:\*  
16: em\_fun:\*  
17: em\_hum:\*  
18: em\_in:\*  
19: em\_mu:\*  
20: em\_om:\*  
21: em\_or:\*  
22: em\_ov:\*  
23: em\_pat:\*  
24: em\_ph:\*  
25: em\_pl:\*  
26: em\_ro:\*  
27: em\_sts:\*  
28: em\_un:\*  
29: em\_vl:\*  
30: em\_hgt\_hum:\*  
31: em\_hgt\_inv:\*  
32: em\_hgt\_other:\*  
33: em\_hgt\_inv:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length DB	ID	Description
-----					

1	16	100.0	1172	8	AF041051	AF041051 Populus t
2	16	100.0	4397	8	AY043494	AY043494 Populus t
3	16	100.0	40114	3	AF067942	AF067942 Caenorh
4	16	100.0	173705	2	AC012256	AC012256 Homo sapi
5	16	100.0	177050	10	AC084390	AC084390 Mus muscu
6	16	100.0	221647	10	AL591003	AL591003 Mouse DNA
7	15	93.8	3545	8	SCYOL075C	274817 S.cerevisia
8	15	93.8	4423	1	HEP06609	U86609 Helicobact
9	15	93.8	17680	1	AE000625	AE000625 Helicobac
10	15	93.8	26236	2	AC107593	AC107593 Rattus no
11	15	93.8	26458	2	AC005942	AC005942 Homo sapi
12	15	93.8	30208	9	HS212J10	AL034401 Human DNA
13	15	93.8	46363	2	AC099662	AC099662 Rattus no
14	15	93.8	62490	2	AC110265	AC110265 Mus muscu
15	15	93.8	76122	2	AC098003	AC098003 Rattus no
16	15	93.8	78379	8	AB006696	AB006696 Arabidops
17	15	93.8	84592	2	AC016403	AC016403 Homo sapi
18	15	93.8	91894	2	AC005739	AC005739 Homo sapi
19	15	93.8	92134	2	AC018406	AC018406 Homo sapi
20	15	93.8	94908	2	AC099247	AC099247 Rattus no
21	15	93.8	126062	2	AC011363	AC011363 Homo sapi
22	15	93.8	131312	2	AC008456	AC008456 Homo sapi
23	15	93.8	133563	2	AC099481	AC099481 Homo sapi
24	15	93.8	133589	2	AC103534	AC103534 Rattus no
25	15	93.8	134078	9	AC068515	AC068515 Homo sapi
26	15	93.8	137330	2	AC092346	AC092346 Homo sapi
27	15	93.8	150157	2	AC103150	AC103150 Rattus no
28	15	93.8	150288	2	AC015572	AC015572 Homo sapi
29	15	93.8	150288	2	AC015574	AC015574 Homo sapi
30	15	93.8	151872	2	AC084250	AC084250 Homo sapi
31	15	93.8	152743	2	AC103511	AC103511 Rattus no
32	15	93.8	153647	2	AC079992	AC079992 Homo sapi
33	15	93.8	157487	2	AC095944	AC095944 Rattus no
34	15	93.8	157656	2	AC104124	AC104124 Homo sapi
35	15	93.8	158001	2	AC104365	AC104365 Homo sapi
36	15	93.8	159150	2	AC022863	AC022863 Homo sapi
37	15	93.8	161775	9	AC099512	AC099512 Homo sapi
38	15	93.8	164958	2	AC094338	AC094338 Rattus no
39	15	93.8	167438	2	AC095832	AC095832 Rattus no
40	15	93.8	171949	9	AC012531	AC012531 Homo sapi
41	15	93.8	172403	2	AC017061	AC017061 Homo sapi
42	15	93.8	177512	2	AC099364	AC099364 Rattus no
43	15	93.8	177968	30	AC022736	AC022736 Homo sapi
44	15	93.8	180638	9	AC068763	AC068763 Homo sapi
45	15	93.8	181127	2	AC068203	AC068203 Homo sapi

## ALIGNMENTS

RESULT	1	AF041051	LOCUS	DEFINITION	ACCESSION	VERSION	KEYWORDS	SOURCE	ORGANISM	REFERENCE	TITLE	JOURNAL	MEDLINE	REFERENCE
AF041051	1172 bp	DNA	linear	Populus tremuloides clone pL4C1p 4-commarate:CoA ligase gene,	AF041051				Populus tremuloides	1	quaking aspen.			
AF041051	1172 bp	DNA	linear	Populus tremuloides clone pL4C1p 4-commarate:CoA ligase gene,	AF041051				Populus tremuloides	2	quaking aspen.			
AF041051	1172 bp	DNA	linear	Populus tremuloides clone pL4C1p 4-commarate:CoA ligase gene,	AF041051				Populus tremuloides	3	quaking aspen.			
AF041051	1172 bp	DNA	linear	Populus tremuloides clone pL4C1p 4-commarate:CoA ligase gene,	AF041051				Populus tremuloides	4	quaking aspen.			
AF041051	1172 bp	DNA	linear	Populus tremuloides clone pL4C1p 4-commarate:CoA ligase gene,	AF041051				Populus tremuloides	5	quaking aspen.			
AF041051	1172 bp	DNA	linear	Populus tremuloides clone pL4C1p 4-commarate:CoA ligase gene,	AF041051				Populus tremuloides	6	quaking aspen.			
AF041051	1172 bp	DNA	linear	Populus tremuloides clone pL4C1p 4-commarate:CoA ligase gene,	AF041051				Populus tremuloides	7	quaking aspen.			
AF041051	1172 bp	DNA	linear	Populus tremuloides clone pL4C1p 4-commarate:CoA ligase gene,	AF041051				Populus tremuloides	8	quaking aspen.			
AF041051	1172 bp	DNA	linear	Populus tremuloides clone pL4C1p 4-commarate:CoA ligase gene,	AF041051				Populus tremuloides	9	quaking aspen.			
AF041051	1172 bp	DNA	linear	Populus tremuloides clone pL4C1p 4-commarate:CoA ligase gene,	AF041051				Populus tremuloides	10	quaking aspen.			
AF041051	1172 bp	DNA	linear	Populus tremuloides clone pL4C1p 4-commarate:CoA ligase gene,	AF041051				Populus tremuloides	11	quaking aspen.			
AF041051	1172 bp	DNA	linear	Populus tremuloides clone pL4C1p 4-commarate:CoA ligase gene,	AF041051				Populus tremuloides	12	quaking aspen.			
AF041051	1172 bp	DNA	linear	Populus tremuloides clone pL4C1p 4-commarate:CoA ligase gene,	AF041051				Populus tremuloides	13	quaking aspen.			
AF041051	1172 bp	DNA	linear	Populus tremuloides clone pL4C1p 4-commarate:CoA ligase gene,	AF041051				Populus tremuloides	14	quaking aspen.			
AF041051	1172 bp	DNA	linear	Populus tremuloides clone pL4C1p 4-commarate:CoA ligase gene,	AF041051				Populus tremuloides	15	quaking aspen.			
AF041051	1172 bp	DNA	linear	Populus tremuloides clone pL4C1p 4-commarate:CoA ligase gene,	AF041051				Populus tremuloides	16	quaking aspen.			
AF041051	1172 bp	DNA	linear	Populus tremuloides clone pL4C1p 4-commarate:CoA ligase gene,	AF041051				Populus tremuloides	17	quaking aspen.			
AF041051	1172 bp	DNA	linear	Populus tremuloides clone pL4C1p 4-commarate:CoA ligase gene,	AF041051				Populus tremuloides	18	quaking aspen.			
AF041051	1172 bp	DNA	linear	Populus tremuloides clone pL4C1p 4-commarate:CoA ligase gene,	AF041051				Populus tremuloides	19	quaking aspen.			
AF041051	1172 bp	DNA	linear	Populus tremuloides clone pL4C1p 4-commarate:CoA ligase gene,	AF041051				Populus tremuloides	20	quaking aspen.			
AF041051	1172 bp	DNA	linear	Populus tremuloides clone pL4C1p 4-commarate:CoA ligase gene,	AF041051				Populus tremuloides	21	quaking aspen.			
AF041051	1172 bp	DNA	linear	Populus tremuloides clone pL4C1p 4-commarate:CoA ligase gene,	AF041051				Populus tremuloides	22	quaking aspen.			
AF041051	1172 bp	DNA	linear	Populus tremuloides clone pL4C1p 4-commarate:CoA ligase gene,	AF041051				Populus tremuloides	23	quaking aspen.			
AF041051	1172 bp	DNA	linear	Populus tremuloides clone pL4C1p 4-commarate:CoA ligase gene,	AF041051				Populus tremuloides	24	quaking aspen.			
AF041051	1172 bp	DNA	linear	Populus tremuloides clone pL4C1p 4-commarate:CoA ligase gene,	AF041051				Populus tremuloides	25	quaking aspen.			
AF041051	1172 bp	DNA	linear	Populus tremuloides clone pL4C1p 4-commarate:CoA ligase gene,	AF041051				Populus tremuloides	26	quaking aspen.			
AF041051	1172 bp	DNA	linear	Populus tremuloides clone pL4C1p 4-commarate:CoA ligase gene,	AF041051				Populus tremuloides	27	quaking aspen.			
AF041051	1172 bp	DNA	linear	Populus tremuloides clone pL4C1p 4-commarate:CoA ligase gene,	AF041051				Populus tremuloides	28	quaking aspen.			
AF041051	1172 bp	DNA	linear	Populus tremuloides clone pL4C1p 4-commarate:CoA ligase gene,	AF041051				Populus tremuloides	29	quaking aspen.			
AF041051	1172 bp	DNA	linear	Populus tremuloides clone pL4C1p 4-commarate:CoA ligase gene,	AF041051				Populus tremuloides	30	quaking aspen.			
AF041051	1172 bp	DNA	linear	Populus tremuloides clone pL4C1p 4-commarate:CoA ligase gene,	AF041051				Populus tremuloides	31	quaking aspen.			
AF041051	1172 bp	DNA	linear	Populus tremuloides clone pL4C1p 4-commarate:CoA ligase gene,	AF041051				Populus tremuloides	32	quaking aspen.			
AF041051	1172 bp	DNA	linear	Populus tremuloides clone pL4C1p 4-commarate:CoA ligase gene,	AF041051				Populus tremuloides	33	quaking aspen.			
AF041051	1172 bp	DNA	linear	Populus tremuloides clone pL4C1p 4-commarate:CoA ligase gene,	AF041051				Populus tremuloides	34	quaking aspen.			
AF041051	1172 bp	DNA	linear	Populus tremuloides clone pL4C1p 4-commarate:CoA ligase gene,	AF041051				Populus tremuloides	35	quaking aspen.			
AF041051	1172 bp	DNA	linear	Populus tremuloides clone pL4C1p 4-commarate:CoA ligase gene,	AF041051				Populus tremuloides	36	quaking aspen.			
AF041051	1172 bp	DNA	linear	Populus tremuloides clone pL4C1p 4-commarate:CoA ligase gene,	AF041051				Populus tremuloides	37	quaking aspen.			
AF041051	1172 bp	DNA	linear	Populus tremuloides clone pL4C1p 4-commarate:CoA ligase gene,	AF041051				Populus tremuloides	38	quaking aspen.			
AF041051	1172 bp	DNA	linear	Populus tremuloides clone pL4C1p 4-commarate:CoA ligase gene,	AF041051				Populus tremuloides	39	quaking aspen.			
AF041051	1172 bp	DNA	linear	Populus tremuloides clone pL4C1p 4-commarate:CoA ligase gene,	AF041051				Populus tremuloides	40	quaking aspen.			
AF041051	1172 bp	DNA	linear	Populus tremuloides clone pL4C1p 4-commarate:CoA ligase gene,	AF041051				Populus tremuloides	41	quaking aspen.			
AF041051	1172 bp	DNA	linear	Populus tremuloides clone pL4C1p 4-commarate:CoA ligase gene,	AF041051				Populus tremuloides	42	quaking aspen.			
AF041051	1172 bp	DNA	linear	Populus tremuloides clone pL4C1p 4-commarate:CoA ligase gene,	AF041051				Populus tremuloides	43	quaking aspen.			
AF041051	1172 bp	DNA	linear	Populus tremuloides clone pL4C1p 4-commarate:CoA ligase gene,	AF041051				Populus tremuloides	44	quaking aspen.			
AF041051	1172 bp	DNA	linear	Populus tremuloides clone pL4C1p 4-commarate:CoA ligase gene,	AF041051				Populus tremuloides	45	quaking aspen.			

TITLE Direct Submission  
JOURNAL Submitted (06-JAN-1998) Plant Biotechnology Research Center, School of Forestry and Wood Products, Michigan Technological University, 1400 Townsend Drive, Houghton, MI 49931, USA

FEATURES  
source  
1. .1172  
/organism="Populus tremuloides"  
/db\_xref="taxon:3693"  
/clone="Pt4CLP"  
/tissue.type="leaf"  
1. .>1172  
/gene="4-coumarate:CoA ligase"  
1. .1172  
/gene="4-coumarate:CoA ligase"  
BASE COUNT 399 a 224 c 181 g 368 t  
ORIGIN

Query Match 100.0%; Score 16; DB 8; Length 1172;  
Best Local Similarity 100.0%; Pred. No. 90;  
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ccttcacacacccc 16  
Db 1018 CCTTCACCAACCCC 1033

RESULT 2  
AY043494 4397 bp DNA linear PLN 17-SEP-2001  
DEFINITION Populus tomentosa 4-coumarate:CoA ligase gene, complete cds.  
ACCESSION AY043494  
VERSION AY043494.1 GI:15636676  
KEYWORDS  
SOURCE Populus tomentosa.  
ORGANISM Populus tomentosa.  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids I; Malpighiales; Salicaceae; Populus.  
1 (bases 1 to 4397)  
Lu, H., Zeng, Q. and Jiang, X.  
Genomic DNA sequence of 4-coumarate:CoA ligase gene from Populus tomentosa  
Unpublished  
2 (bases 1 to 4397)  
Lu, H., Zeng, Q. and Jiang, X.  
Direct Submission  
Submitted (01-JUL-2001) Biology, Beijing Forestry University, Beijing Qinghua East Road No. 35, Beijing 100081, China  
Location/Qualifiers  
1. .4397  
/organism="Populus tomentosa"  
/db\_xref="taxon:118781"  
1. .1136  
join(1137..2136,3321..3519,3644..3857,3957..4059,4174..4397)  
/product="4-coumarate:CoA ligase"  
join(1153..2136,3321..3519,3644..3857,3957..4059,4174..4284)  
/codon\_start=1  
/product="4-coumarate:CoA ligase"  
/protein\_id="AL02144.1"  
/db\_xref="GI:15636677"  
/translation="NNROEFIRSKLPIPIYIPKNLPLSHVLENTSNHSKPCILINGANGDYVATDELTAARVASGLKIGIOGDVIMELPSSPEVLAFLGASHGALITANPESTPAELAKHAKASRAKELITIOACTYERKVDARSADVAVKCDAPDCLEHSELTQADENARPQVDSPPDVVALPYSSGTGLPKGYMLTHKGLITSVAQVDNDNLYVSEVDNLCVLPFHIVILNSIMLCGLRVGASILIMPKFEISGLILEKYSIAPVVPVMAISAKSPDLKHDLSLRMKSGAPLGKLEDTVRAKFPQARGLGQYGTAGPVLAACLAFAKEPDKPGAGCTVYRNAEMKIVDPENGASLRPMQPGICRGQIMKGYLNDPRATSKTIDKEGMHIGDYGITDDDELFTVDRKLKILYKKGQVAPALEALIAHPEISDAVAVGLKDEDAGEVAVFVAVSEKQATEDETKOYISKOVIFPKRIKRVFFEAIAPKAPSGKILRKLEKLAGI"

FEATURES  
source  
1. .4397  
/organism="Populus tomentosa"  
/db\_xref="taxon:118781"  
1. .1136  
join(1137..2136,3321..3519,3644..3857,3957..4059,4174..4397)  
/product="4-coumarate:CoA ligase"  
join(1153..2136,3321..3519,3644..3857,3957..4059,4174..4284)  
/codon\_start=1  
/product="4-coumarate:CoA ligase"  
/protein\_id="AL02144.1"  
/db\_xref="GI:15636677"  
/translation="NNROEFIRSKLPIPIYIPKNLPLSHVLENTSNHSKPCILINGANGDYVATDELTAARVASGLKIGIOGDVIMELPSSPEVLAFLGASHGALITANPESTPAELAKHAKASRAKELITIOACTYERKVDARSADVAVKCDAPDCLEHSELTQADENARPQVDSPPDVVALPYSSGTGLPKGYMLTHKGLITSVAQVDNDNLYVSEVDNLCVLPFHIVILNSIMLCGLRVGASILIMPKFEISGLILEKYSIAPVVPVMAISAKSPDLKHDLSLRMKSGAPLGKLEDTVRAKFPQARGLGQYGTAGPVLAACLAFAKEPDKPGAGCTVYRNAEMKIVDPENGASLRPMQPGICRGQIMKGYLNDPRATSKTIDKEGMHIGDYGITDDDELFTVDRKLKILYKKGQVAPALEALIAHPEISDAVAVGLKDEDAGEVAVFVAVSEKQATEDETKOYISKOVIFPKRIKRVFFEAIAPKAPSGKILRKLEKLAGI"

BASE COUNT 1394 a 823 c 819 g 1361 t  
ORIGIN

Query Match 100.0%; Score 16; DB 8; Length 4397;  
Best Local Similarity 100.0%; Pred. No. 95;  
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ccttcacacacccc 16  
Db 1021 CCTTCACCAACCCC 1036

RESULT 3  
AF067942 40114 bp DNA linear INV 05-OCT-2001  
DEFINITION Caenorhabditis elegans cosmid ZK6, complete sequence.  
ACCESSION AF067942  
VERSION AF067942.1 GI:3165573  
KEYWORDS HTG.  
SOURCE Caenorhabditis elegans.  
ORGANISM Caenorhabditis elegans.  
Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.  
1 (bases 1 to 40114)  
The C. elegans Sequencing Consortium.  
Genome sequence of the nematode C. elegans: a platform for investigating biology. The C. elegans Sequencing Consortium Science 282 (5396), 2012-2018 (1998)  
99069613  
2 (bases 1 to 40114)  
Wu, X.  
The sequence of C. elegans cosmid ZK6  
Unpublished  
3 (bases 1 to 40114)  
Waterston, R.  
Direct Submission  
Unpublished  
4 (bases 1 to 40114)  
Waterston, R.  
Direct Submission  
Submitted (23-MAY-1998) Department of Genetics, Washington University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA  
5 (bases 1 to 40114)  
Waterston, R.  
Direct Submission  
Submitted (03-JAN-2001) Department of Genetics, Washington University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA  
6 (bases 1 to 40114)  
Waterston, R.  
Direct Submission  
Submitted (25-APR-2001) Department of Genetics, Washington University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA  
7 (bases 1 to 40114)  
Waterston, R.  
Direct Submission  
Submitted (07-SEP-2001) Department of Genetics, Washington University, Genome Sequencing Center, 4444 Forest Park Avenue, St. Louis, MO 63110, USA  
8 (bases 1 to 40114)  
Waterston, R.  
Direct Submission  
Submitted (05-OCT-2001) Department of Genetics, Washington University, Genome Sequencing Center, 4444 Forest Park Avenue, St. Louis, MO 63110, USA  
COMMENT  
Genome Sequencing Center  
Department of Genetics, Washington University  
St. Louis, MO 63110, USA, and  
Sanger Centre, Hinxton Hall  
Cambridge CB10 1RQ, England  
email: tw@nematode.wustl.edu and jes@sanger.ac.uk

NOTICE: This sequence may not be the entire insert of this clone. It may be shorter because we only sequence overlapping sections once, or longer because we provide a small overlap between neighboring submissions.

This sequence was finished as follows unless otherwise noted: all regions were double stranded, sequenced with an alternate chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one ml3 subclone.

## NEIGHBORING COSMID INFORMATION

The 5' cosmid is Y39DBB; 3' cosmid is Y39DBA. Actual start of this cosmid is at base position 1 of CELZK6; actual end is at 40114 of CELZK6. The orientation of this cosmid is unknown.

## NOTES:

Coding sequences below are predicted from computer analysis, using the program GeneFinder (P. Green and L. Hillier, ms in preparation).

FEATURES  
source

Location/Qualifiers

1..40114

/organism="Caenorhabditis elegans"

/strain="Bristol N2"

/db\_xref="taxon:6239"

/chromosome="V"

/clone="ZK6"

2105..4708

/gene="ZK6.8"

join(2105..2307,3036..3223,3287..3515,3586..3905,3956..4127,4306..4441,4490..4708)

gene  
CDS

/gene="ZK6.8"

/codon\_start=1

/product="Hypothetical protein ZK6.8"

/protein\_id="AAG45570.1"

/db\_xref="GI:12019644"

/translation="MSEELPPSSGSKMSPRAFLCVMYFGHLICIMGDSQAFI  
LESYINHERDPAKINSHAGYGOATCYLAFTCLSPFLATSAKTLITIAAC  
FTSPFLFLTYNORYTFPSALNGVGRALYTGNGGLTSHSTRKTESNSISVAIG  
SSCMIVAGIILITLTFAGSGSEVADLANAVTOHERESDTEIYLLFVGAHS  
FVGCITALLPSNDIGNCIESKTIYAFRDIAMVAFRSPKMIYVLPVFLVGH  
SFWSIYPTLTFNSHLSAMLYPAISFVGGLGETMGLISFCSKRIKFGMPM  
FICGLICICALVVISPTAPAPAPSEKPLFOPRHYLVIFLIALIGMSDCCISY  
RSVYCATMPRRNQASVSKFOAIGCCVFFTSPLNITYYIGIPILCIITASVC  
FEETRRIKOMEKSLTNMELDQAOORRSSKYDTLDEE"

gene

/gene="ZK6.7"

join(5937..6023,6070..6169,6669..7138,7228..7527,7589..7687)

/gene="ZK6.7"

/note="coded for by the following C. elegans cDNAs:

YK788c11.3, YK788c11.5"

/codon\_start=1

/product="Hypothetical protein ZK6.7b"

/protein\_id="AAL11486.1"

/db\_xref="GI:15967144"

CDS

/translation="MMRPAYFLAEPVODVGSNGDEPLHMTTPQIIEKGYPMIYI  
VAIDGTYILEMHRIPFGKTNVTPNGKRPVVEFMQGLICASSDWVNLPPDSAGFLA  
DAGDVMGLNMGNTYSMKHKDLPSHSAFMDMSDEMATYDLNAMIHNVLEVGD  
VYVYMGSGTLTFNSHLSKDGSPAKIKKFFALPISVYKHIGFSLFANYSLE  
DGMFDIGAGEFLTFNNMAMKLAADICGGLKVEADLCNVLEFLIAGPESDMONOTRP  
VYATHPDAGSTONIVHMQVNHGVPAYWGTRKTKKKYGGSTLPLSREPTST  
GVMPICMLIRLTCTTY"  
join(5937..6023,6070..6169,6669..7138,7228..7527,7576..7713,7765..7881)

/gene="ZK6.7"

/note="similar to lysosomal acid lipases (SW:P38571);  
coded for by the following C. elegans cDNAs: YK63b10.3,  
YK25b10.5, YK308a8.5, YK552a3.3, YK552a3.5, YK222d9.5,  
YK722d9.3, YK73a11.5, YK73g10.5, YK736f6.5, YK736f6.3,  
YK73g10.3, YK851d01.3, YK851d01.5"

/codon\_start=1  
/product="Hypothetical protein ZK6.7a"  
/protein\_id="AAG45574.1"  
/db\_xref="GI:12019648"

/translation="MMRPAYFLAEPVODVGSNGDEPLHMTTPQIIEKGYPMIYI  
VAIDGTYILEMHRIPFGKTNVTPNGKRPVVEFMQGLICASSDWVNLPPDSAGFLA  
DAGDVMGLNMGNTYSMKHKDLPSHSAFMDMSDEMATYDLNAMIHNVLEVGD  
VYVYMGSGTLTFNSHLSKDGSPAKIKKFFALPISVYKHIGFSLFANYSLE  
DGMFDIGAGEFLTFNNMAMKLAADICGGLKVEADLCNVLEFLIAGPESDMONOTRP  
VYATHPDAGSTONIVHMQVNHGVPAYWGTRKTKKKYGGSTLPLSREPTST  
GVMPICMLIRLTCTTY"  
IKLCTDYLGK"  
9204..11716

gene

/gene="ZK6.6"

join(9204..9343,9554..9741,9787..9883,9943..10074,10127..10257,10841..11029,11081..11256,11314..11352,11501..11716)

gene

/gene="ZK6.6"

CDS

/gene="ZK6.6"

/codon\_start=1

/product="Hypothetical protein ZK6.6"

/protein\_id="AAG45573.1"

/db\_xref="GI:12019647"

/translation="MDKFCDETFYKLLKKSYSYKILFLENPAKLRNENVYTYC  
DLXQAVCYLAVVTCLEFSPFLVATSAKTYLLISSICEPSPLGLFTNSYVYFSSA  
LNGVAFALYTGNGGYITSHSTRQTIENSVLSMFCGCCMIVGSIAMVATVYODQ  
IPOLIDAVNTVIGGVERRRDEIYLLFVFAAISMAIFTFMLMPSODVNCIEP  
SDKYVSPFGDKMAETLKSMMKRIAPPTLLTGMVAFWVSIYPTSLTFMNSKMI  
YLPAIYGVGVGETINGIITSLSKRKDKGICPTMIGLSPVLTTRCFVALLSPD  
ATVPSHOPLPFRPRTSISDCNNRNSLSCFTFSLPFLNHYVIGIPILCY  
LSCILFFQAROTQVMERKILQELEESERRMARLEAOMOKI"  
12832..14865

gene

/gene="ZK6.5"

CDS

/gene="ZK6.5"

join(12832..12894,13721..13817,13864..14452,14605..14686,14731..14865)

gene

/gene="ZK6.5"

/note="contains similarity to C4-type zinc finger (Pfam:

zf-C4.hmm, score: 37.95)"

/codon\_start=1

/product="Hypothetical protein ZK6.5"

/protein\_id="AAG45580.1"

/db\_xref="GI:12019654"

gene

/gene="ZK6.4"

join(17012..17074,17511..17607,17658..18272,18679..18941)

CDS

/gene="ZK6.4"

/note="contains similarity to C4-type zinc finger (Pfam:

zf-C4.hmm, score: 34.58) and to the ligand-binding domain

of nuclear hormone receptors (Pfam: hormone\_rec.hmm,

score: 38.56)"

/codon\_start=1

/product="Hypothetical protein ZK6.4"

/protein\_id="AAG45579.1"

/db\_xref="GI:12019653"

gene

/translation="WATNOCVANCERTTEPENGVCSCNACKIFRRLLITRTTPYK  
CYLGNCTCTPTTQKCVCFRQKCTIOGMPLPSLHGEQTKECIDSTIONMLLD  
AHRBIDLNAVNTSYLDPTIDVIRLNKVDYIRKSGSHQINPYNAFHCISLVYDPMK  
FPFVLLFEEQOKYLLKPEYIKLIVFISKSKMSGKCMFPDGSVLPASASMI  
SKISPNLEKIRCRILGRISERITDEYLLMGCYLIFCPNPSLSQISENGRLLTYOH  
LXSSALLOXYCLITYEKAPARFAELLGICHVIGVYDAIHYNVLVQLTKARLDVKOL  
IKDGDAAFNCN"  
20442..21787

CDS

/gene="ZK6.3"

join(20442..20615,20667..20801,21151..21222,21444..21631,

Query Match 100.0%; Score 16; DB 3; Length 40114;  
Best Local Similarity 100.0%; Pred. No. 1e+02;  
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;





```

misc_feature      136736..152397
                  /note="assembly_fragment"
misc_feature      152498..173172
                  /note="assembly_fragment"
misc_feature      173273..173705
                  /note="assembly_fragment"
                  clone_end:77
                  vector_side:right"
BASE COUNT      54075 a 32056 c 33141 g 52665 t 1768 others
ORIGIN

Query Match      100.0%; Score 16; DB 2; Length 173705;
Best Local Similarity 100.0%; Pred. No. 1.1e+02;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      1 ccttcaccaacccc 16
        |||||||
Db      18624 CCTTCACCAACCCC 18609

RESULT 5
AC084390      AC084390      177050 bp      DNA      linear      ROD 26-SEP-2001
LOCUS
DEFINITION    Mus musculus clone RP23-285H9, complete sequence.
ACCESSION     AC084390
VERSION       AC084390.1 GI:15778823
KEYWORDS      HTG.
SOURCE        house mouse.
ORGANISM      Mus musculus
               Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
               Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus;
REFERENCE     1 (bases 1 to 177050)
AUTHORS       McPherson,J.D.
TITLE         The sequence of Mus musculus clone
JOURNAL       Unpublished
REFERENCE     2 (bases 1 to 177050)
AUTHORS       McPherson,J.D.
TITLE         Direct Submission
JOURNAL       Submitted (26-SEP-2001) Genetics, Genome Sequencing Center, 4444
               Forest Park Parkway, St. Louis, Missouri 63108, USA
COMMENT       Center Project name: M_BA0285H09.
FEATURES
             source
               1..177050
               /organism="Mus musculus"
               /db_xref="taxon:10090"
               /clone="RP23-285H9"
BASE COUNT     47756 a 40786 c 41518 g 46990 t
ORIGIN

Query Match      100.0%; Score 16; DB 10; Length 177050;
Best Local Similarity 100.0%; Pred. No. 1.1e+02;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      1 ccttcaccaacccc 16
        |||||||
Db      17118 CCTTCACCAACCCC 17133

RESULT 6
AL591003      AL591003      221647 bp      DNA      linear      ROD 29-NOV-2001
LOCUS
DEFINITION    Mouse DNA sequence from clone RP23-354I24 on chromosome 13,
               complete sequence.
ACCESSION     AL591003
VERSION       AL591003.16 GI:16605732
KEYWORDS      HTG.
SOURCE        house mouse.
ORGANISM      Mus musculus
               Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
               Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

```

```

REFERENCE       1 (bases 1 to 221647)
AUTHORS         Kay,M.
TITLE           Direct Submission
JOURNAL         Submitted (29-NOV-2001) Wellcome Trust Sanger Institute, Hinxton,
               Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
               humquerry@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
               On Nov 3, 2001 this sequence version replaced gi:16030192.
               During sequence assembly data is compared from overlapping clones.
               Where differences are found these are annotated as variations
               together with a note of the overlapping clone name. Note that the
               variation annotation may not be found in the sequence submission
               corresponding to the overlapping clone, as we submit sequences with
               only a small overlap as described above.
               This sequence was finished as follows unless otherwise noted: all
               regions were either double-stranded or sequenced with an alternate
               chemistry or covered by high quality data (i.e., phred quality >=
               30); an attempt was made to resolve all sequencing problems, such
               as compressions and repeats; all regions were covered by at least
               one plasmid subclone or more than one M13 subclone; and the
               assembly was confirmed by restriction digest. The following
               abbreviations are used to associate primary accession numbers given
               in the feature table with their source databases: Em: EMBL; Sw:
               SWISSPROT; Tr: TrEMBL; Wp: WormPEP; Information on the WormPEP
               database can be found at
               http://www.sanger.ac.uk/projects/Celegans/wormpep RP23-354I24 is
               from the RPI-23 Mouse PAC library
               constructed by the group of Pieter de Jong.
               For further details see http://www.chori.org/bacpac/home.htm
               VECTOR: pBACE3.6
               This sequence is the entire insert of clone RP23-354I24.
FEATURES
             source
               1..221647
               /organism="Mus musculus"
               /db_xref="taxon:10090"
               /chromosome="13"
               /clone="RP23-354I24"
               /clone_1lb="RPI-23"
               /note="Single clone region. Sequence from reads from a
               short insert library derived from a single pUC clone.
               Restriction digest data confirm the assembly."
misc_feature
Qy      1 ccttcaccaacccc 16
        |||||||
Db      152713 CCTTCACCAACCCC 152698

Query Match      100.0%; Score 16; DB 10; Length 221647;
Best Local Similarity 100.0%; Pred. No. 1.1e+02;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      1 ccttcaccaacccc 16
        |||||||
Db      152713 CCTTCACCAACCCC 152698

RESULT 7
SCYOL075C/c     SCYOL075C      3545 bp      DNA      linear      PLN 05-AUG-1997
LOCUS
DEFINITION    S.cerevisiae chromosome XV reading frame ORF YOL075c.
ACCESSION     Z74817 Y13140
VERSION       Z74817.1 GI:1419904
KEYWORDS
SOURCE        baker's yeast.
ORGANISM      Saccharomyces cerevisiae
               Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
               Saccharomycetales; Saccharomycetaceae; Saccharomyces.
REFERENCE     1 (bases 1 to 3545)
AUTHORS       Alexandrakl,D., Katsoulou,C. and Tzermitia,M.
JOURNAL       Unpublished
REFERENCE     2 (bases 1 to 3545)
AUTHORS       MIPS.
TITLE         Direct Submission
JOURNAL       Submitted (04-JUL-1996) Data collected by MIPS on behalf of the
               European yeast chromosome XV sequencing project. MIPS at the

```

Max-Planck-Institut fuer Biochemie, Am Klopfersplitz 18a D-82152 Martinsried, FRG; E-mail: Mewes@mpi-biochemie.mpg.de

## FEATURES

## SOURCE

1. 3545

/organism="Saccharomyces cerevisiae"

/db\_xref="taxon:4932"

/chromosome="XV"

/complement(244..>3545)

/note="orf y0r075c"

/codon\_start=3

/protein\_id="CAA99085.1"

/db\_xref="GI:2315260"

/translation="IGTOMISNPISIMFLDEPTTGDAVSFLVYKTIKTLAKEDGRTF

LSMSHPSRDLIFLDQVCIISGNVYCDKMDNTIPESIGYHVPOLVNPADYFID

IMSDRSDEKKAATGSRINSLIDHMDYERTHOLQAEYSINATLQIONMTTRP

FMKQVTLTRNFKLNFSDYVTLSTAPELITVCGWIIYKDKSISGLRTTAC

LYASTIITCYITLFTYRLCEODIALDKRAQSVPLAFYARKISLSDDEAM

TMIFVITTFMGLADARKFYQFAVFLQSLGSLMSVAVSRDPSKASLVGM

TFVLSMGCEFNAAKVPYVYRMKIAFTWSEGLMSSTFNSTCTDNDECIG

NOILEVYGFPRMTTPAVVLLCWVGVFVGAIIYKIDITLQNEVSKOKIK

KSPGKMEIOLDDVYHOKDLEAKGNHITIKEDIDLRVTFESAPESMKRGNF

HETREILOSVAIEFKPGMINAINGPSSGSSLLNLSGRKSVFAKPTDSSIMN

DIQVSELMFNKVCYVSQDDHLALIVKETLYAALRLDHLTEAKREDNLIIR

SLGKHCENNIIGNEFKVIGSGEKRRVTVGLNDPILLDEPTSGDSEFSATI

LEILKICRQGGKTIITIHOPRESEFKREGNVLLAKSGTAFNGSPDEISATF

GYNCPSTNVADEFLDISYNTQONEISRAVERKLSAMKANMDESISPISE

KOQYSEFTESEFVRKPNALVAVYVNRKROFTTTRSPSLMARIAOIPGLVI

FALEFAPVKNHTYSISNRGLAESTALYFGMGLACYPTEEDYFEEYNVYI

APPELAYMLEPLSLASVLYANFTYLACGPRTPANFRATYCSFTVCCGRLEI

MTNTPFERPGEFVNCISITISIGTOMSLGSLGRVLEKGFNVLPVGYSMIINFA

PGNKLTCDEGKNSDGTCEFANGHVLVSYGRNTORYLGIIVCAIYRLIAEF

ILKAKLEIMW"

BASE COUNT 1156 a 707 c 585 g 1097 t

ORIGIN

Query Match 93.8%; Score 15; DB 8; Length 3545;  
Best Local Similarity 100.0%; Pred. No. 3.9e+02;  
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ccttcacacacccc 15  
Db 1174 CCTTCACCAACCCC 1160

RESULT 8 4423 bp DNA linear BCT 31-AUG-1999  
HP06609/c  
LOCUS  
DEFINITION Helicobacter pylori ribosomal protein L1 homolog gene, partial cds,  
50S ribosomal protein L7/L12 gene, complete cds, and DNA-directed  
RNA polymerase homology gene, partial cds.

ACCESSION 066609.1 GI:1840149

VERSION 066609.1  
KEYWORDS  
SOURCE Helicobacter pylori.  
ORGANISM Helicobacter pylori.  
Bacteria; Proteobacteria; epsilon subdivision; Helicobacter group;  
Helicobacter.

REFERENCE 1 (bases 1 to 4423)  
AUTHORS Hocking, D., Webb, E., Radcliff, F., Rothel, L., Taylor, S.,  
Pincus, G., Kapoulas, C., Braley, H., Lee, A. and Doidge, C.

TITLE Isolation of recombinant protective Helicobacter pylori antigens  
JOURNAL Infect. Immun. 67 (9), 4713-4719 (1999)

PMID 9938687

PMID 10456921

2 (bases 1 to 4423)  
AUTHORS Hocking, D., Rothel, L., Doidge, C., Radcliff, F., Lee, A. and Webb, E.

TITLE Recombinant Helicobacter pylori ribosomal protein of 13kDa mass,  
predicted to be a 50S ribosomal protein L7/L12

JOURNAL Unpublished  
3 (bases 1 to 4423)  
REFERENCE Hocking, D., Rothel, L., Doidge, C., Radcliff, F., Lee, A. and Webb, E.

AUTHORS Direct Submission

JOURNAL Submitted (23-JAN-1997) Microbiology Research, CSL Ltd., 45 Poplar  
Rd, Parkville, Victoria 3052, Australia

## FEATURES

## SOURCE

1. 4423

/organism="Helicobacter pylori"

/strain="Hp921023"

/db\_xref="taxon:210"

<1..538

/note="similar to bacterial 50S ribosomal protein L1"

/product="ribosomal protein L1 homolog"

/protein\_id="AAB47277.1"

/db\_xref="GI:1840150"

/translation="DQNVKGAVALPHGTGRVAVFAKDIKDEAKNAGAVYGGD

LAELIKGRIDFMVATPPMMAVAVGVGRITLCPKMPKGTGYMDAKAVTNK

SGGVNFRVDRKQVHAFIGASPEEKIKENMELVITNLRKPSAKKYIRNAALS

LTSPSVSLDAQELMDIK"

1186..1563

/note="13 kDa protein"

/codon\_start=1

/transl\_table=1

/product="50S ribosomal protein L7/L12"

/protein\_id="AAB47278.1"

/db\_xref="GI:1840151"

/translation="MAISKEVLEYIGSLVLSLSELYKVFEEKGVSAFTYVAGAA

VAGAAAESEKREFTNITLDSGAEKIKVAVREITGLKAEKADTEKTPHAKEG

VNEEATIKKLEVEAKVEK"

1783..>4423

/note="similar to C. jejuni DNA-directed RNA polymerase

and to E. coli DNA-directed RNA polymerase beta chain"

/codon\_start=1

/transl\_table=1

/product="DNA-directed RNA polymerase homolog"

/protein\_id="AAB47279.1"

/db\_xref="GI:1840152"

/translation="MSKIPILKRLRADFTPTDLEVPNLLLDSDYSDFLKSOG

KESGIRKVFESIPIDENHRTILEVAGCEFGSKTVFRAMRGITYSIPKIKYRL

ILMEKTRSGEKNIGIDIKOSIFIREIPEMPTSGITINGVRVYNOLHSPGYIF

KREESSTLNKLITYGQIIPDRSMITFEYDSVDYLARKIRKRPVITLFRAMDTQ

KODIKMEFLVAVRYENDYLIPLFASLDNQMEFLDQDQKILLAGKILTSRKI

KELKENHLEWVEYPMOILNRLHAEVPMVGKVLMDLQDQNKLEKIHDLGVOEF

IINDLALGHDAISIIHSFLADYBSLAKOTERIDDENALAAIIRHKVMRGDVTVEY

AKQFVKLFEPEDRYDLTWVGRKMNKIGLHVDPYITLTHEDITTYVLYMKIN

OGKIDDRDLGNRIKAVGELIANELHSGIVKMKITIKKITMSGAPDSIMPHDLYN

SKWITSTIMEFEFGOLSOEMOTNPLSEYTHKRLSALGEGLYDRGFEKRDYHP

THYGRICPIETPEBQNTIGLNTLSTFRVNDLGELEAPYKRVVDKRVGEETVLAIO

EDSHIIPASTPIDEEGNIGLDTETRVGEIYELANKSVTLMDLSSMLGVASLI

PLEHDANRKLMTGMORAVPLRLSDAPVGTGIEKIIARDSGAIIANAGAVYER

IDSKNYIIGEGKEAYIDAYSLQKMLRNQNTSPFQVYKVGDKVEAGQIILADPS

MDGELALGKNVAVPMPNGYFEDAIYVSEITRDIDFTSTHYIYEREVDRELKHS

VEEFTADIDIVKEALAHDESGI"

BASE COUNT 1395 a 786 c 1057 g 1185 t

## ORIGIN

Query Match 93.8%; Score 15; DB 1; Length 4423;  
Best Local Similarity 100.0%; Pred. No. 4e+02;  
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2 ccttcacacacccc 16  
Db 3445 CCTTCACCAACCCC 3431

RESULT 9 17880 bp DNA linear BCT 06-APR-1999  
AE000625  
LOCUS

DEFINITION Helicobacter pylori 26695 section 103 of 134 of the complete  
genome.

ACCESSION AE000625 AE000511

VERSION AE000625.1 GI:2314349

KEYWORDS



```

/note="similar to GB:X00415 SP:P02996 GB:J01689 GB:V00356
PID:147895 percent identity: 67.48; identified by sequence
similarity: putative"
/codon_start=1
/transl_table=11
/product="translation elongation factor Ef-G (fusA)"
/protein_id="AD08239.1"
/db_xref="GI:2314354"
/translation="MARKPLNTRINIGIAHIDAGKTTTSERILEFTGVSHKIGEVH
DGAATMDMMOEKRGITTSAAFTCCMKKHQINLIDTPCHVDTTIEVESHMRYLDA
VSVFCYGVQPOSEYVWROANKYGVPRIVFNKMDRIGANFTVENQILRLKANV
PINPIGAEDTFTGVIDLVOMKAIWMNETMGAYDEELPSDLLEKAYREKLVA
VAQDDEALMEKYVGEELSTIEIKKACLNMSIVPMLGSGSEFKKQVOTLDAVI
DYPAPLEVVDIKIDPKEEVEFKSSDGEFAGLAFTMTDPVQGLNFVRYRCK
LESQSVYVNSTKDKKERVGRLLKMSNRKREDIEVAGETICAPVGLKDTLGDPLDDE
KNAVLERMERPEPPIHIAVEPKTKADQEMGVALGLAEDDSFRMTQDEEGQTLI
GGMEHLLEITVDRLKREFVEADIGQVAFRETTIRSSVSKHEKTIKAGSGKQVGH
VFITKLEKPEKSGEYEVNLEISGVITPEKXIPAVDKGIOEMGVGLAYGVVDKFTL
YDGSYHDVDSSEMAFKIAGSAFKEASRANPVLLEPMMEVEVEPEYMGVDYGLDN
RRGQINSMDDRGLKIVNAFVPLVEMFGYSTDLRSATQCRGYYSMEFDHYGEVPSNI
AKIVEKRRG"
gene
complement(8089..8556)
/gene="Hpi196"
complement(8089..8556)
/gene="Hpi196"
/note="similar to SP:P22744 percent identity: 61.44;
identified by sequence similarity: putative"
/codon_start=1
/transl_table=11
/product="ribosomal protein S7 (rps7)"
/protein_id="AD08240.1"
/db_xref="GI:2314355"
/translation="MRRRKAPREVLCDPYGNKVVYKFINKMWDPKKSVAEKTIK
AFKRIEKGSEKGEYERKALERVRLPEYRSRGATQVPEVEVSAOSIETSM
ILATRRNRMMVDRIANELMDAASDKGAFFKKEDVHMAENKFAIYRW"
gene
complement(8572..8979)
/gene="Hpi197"
complement(8572..8979)
/gene="Hpi197"
/note="similar to SP:P18662 percent identity: 79.03;
identified by sequence similarity: putative"
/codon_start=1
/transl_table=11
/product="ribosomal protein s12 (rps12)"
/protein_id="AD08241.1"
/db_xref="GI:2314356"
/translation="MPTINILIRKRRKVVYKTKSPALVCEQRCQVCTVYTTTPK
PNSALRKVAVRLTSKFEVSTYIPGESHNIQENHSIVLVREGRYADLPGRVTHVIRGAL

```

Query Match 93.8%; Score 15; DB 1; Length 17880;

Best Local Similarity 100.0%; Pred. No. 4.2e+02;

Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 cttcaccaccccc 16

Db 16078 CTTTACACCAACCCC 16092

```

RESULT 10
AC107593 26236 bp DNA linear HTG 23-JAN-2002
LOCUS Rattus norvegicus clone CH230-18605, *** SEQUENCING IN PROGRESS
DEFINITION *** 17 unordered pieces.
AC107593 AC107593.1 GI:1826586
VERSION HTG; HTGS_PHASE1.
KEYWORDS Norway rat.
SOURCE Rattus norvegicus
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
REFERENCE 1 (pages 1 to 26236)
AUTHORS Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C.,

```

TITLE  
JOURNAL  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
COMMENT

Unpublished

2 (bases 1 to 26236)

Submitted Submission

Direct Submission

Submitted (23-JAN-2002) Human Genome Sequencing Center, Department

of Molecular and Human Genetics, Baylor College of Medicine, One

Baylor Plaza, Houston, TX 77030, USA

Center: Genome Center

Center: Baylor College of Medicine

Center code: BCM

Web site: <http://www.hgsc.bcm.tmc.edu/>

Contact: [hgsc-help@bcm.tmc.edu](mailto:hgsc-help@bcm.tmc.edu)

Project Information

Center project name: GOVS

Center clone name: CH230-18605

Sequencing Statistics

Chemistry: Dye-terminator Big Dye 98% of reads

Assembly program: Phrap, version 0.990329First call to

findPhrapList

Consensus quality: 21752 bases at least Q40

Consensus quality: 23932 bases at least Q30

Consensus quality: 25882 bases at least Q20

Estimated insert size: 18630; sum-of-contigs estimation

Quality coverage: 0x in Q20 bases; agrose-fp estimation

Quality coverage: 0.1x in Q20 bases; sum-of-contigs estimation

NOTE: Estimated insert size may differ from sequence length

(see [http://www.hgsc.bcm.tmc.edu/docs/genbank/draft\\_data.html](http://www.hgsc.bcm.tmc.edu/docs/genbank/draft_data.html)).

NOTE: This is a "working draft" sequence. It currently

consists of 17 contigs. The true order of the pieces

is not known and their order in this sequence record is

arbitrary. Gaps between the contigs are represented as

\* runs of N, but the exact sizes of the gaps are unknown.

Alsbrooks,S.L., Amaralunga,H.C., Are,J.R., Banks,T., Barbara,J.,

Benton,J., Bimaga,K., Blankenbury,K., Bonin,D., Bouck,J.,

Bowie,S., Brileva,M., Brown,E., Brown,M., Bryant,N.P., Buhay,C.,

Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C., Cartron,T.F.,

Carter,M., Cavazos,S.R., Chacko,J., Chavez,D., Chen,G., Chen,R.,

Chen,Z., Chowdhury,I., Christopoulos,C., Cleveland,C.D., Cox,C.,

Coyle,M.D., Dathorne,S.R., David,R., Davila,M.L., Davis,C.,

Davy-Carroll,L., Dederich,D.A., Delaney,K.R., Delgado,O.,

Denn,A.L., Ding,Y., Dinh,H.H., Douthett,K.J., Draper,H.,

Dugan-Rocha,S., Durbin,K.J., Earnhart,C., Edgar,D., Edwards,C.C.,

Elhaj,C., Escotto,M., Falls,T., Ferraguto,D., Flagg,N., Ford,J.,

Foster,P., Frantz,P., Gabisi,A., Gao,J., Garcia,A., Garner,T.,

Garza,N., Gill,R., Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S.,

Hamilton,K., Harris,C., Harris,K., Hart,M., Havak,P., Hawes,A.,

Hernandez,J., Hernandez,O., Hodson,A., Hogue,M., Holloway,C.,

Hollins,B., Homsi,F., Howard,S., Huber,J., Hulik,S., Hume,J.,

Jackson,L.E., Jacobson,E., Kelly,S., Khan,U., Kling,L., Korvah,J.,

Joudan,S., Karlsson,E., Kelly,S., Khan,U., Kling,L., Korvah,J.,

Kovar,C., Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C.,

Lewis,L., Li,J., Li,Z., Licharge,O., Lien,C., Liu,J., Liu,M.,

Louisege,H., Lozano,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R.,

Ma,J., Maheshwari,M., Mapua,P., Martin,R., Martindale,A.,

Martinez,E., Massey,E., Mawhney,E., McLeod,M.P., Meador,M.,

Mei,G., Metzker,M., Miner,G., Miner,Z., Mitchell,T., Mohabbat,K.,

Morgan,M., Morris,S., Moser,M., Neal,D., Newton,J., Newton,N.,

Nguyen,A., Nguyen,N., Nguyen,N., Nickerson,E., Nokenko,S.,

Oguchi,M., Okunou,G., Oragunye,N., Oyedero,R., Pace,A., Payton,B.,

Peery,J., Perez,L., Peters,L., Pickens,R., Primus,E., Pu,L.L.,

Quiles,M., Ren,Y., Rives,M., Rojas,A., Rojoudokan,I., Rolfe,M.,

Ruliz,S., Saverly,G., Scherer,S., Scott,G., Shen,H., Shoostari,N.,

Slison,H., Sodergren,E., Sonalke,T., Sparks,A., Stanley,H.,

Stone,R., Sutton,A., Svatek,A., Tabors,P., Tamerisa,A., Tamerisa,K.,

Tang,H., Tansey,J., Taylor,C., Taylor,T., Telford,B., Thomas,N.,

Thomas,S., Usmani,K., Vasquez,L., Vera,Y., Villalón,D., Vinson,R.,

Wall,R., Wang,S., Ward-Moore,S., Warren,R., Washington,C.,

Washington,S., Williams,G., Williamson,A., Wiczek,R., Wooden,S.,

Worley,K., Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D.,

Weinstock,G., and Gibbs,R.

\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will  
\* be preserved.

```

* 1 1449: contig of 1449 bp in length
* 1450 1549: gap of unknown length
* 1550 4116: contig of 2567 bp in length
* 4216 4216: gap of unknown length
* 4217 6167: contig of 1951 bp in length
* 6168 6267: gap of unknown length
* 6268 7844: contig of 1577 bp in length
* 7845 7944: gap of unknown length
* 7945 9386: contig of 1442 bp in length
* 9387 9486: gap of unknown length
* 9487 11068: contig of 1582 bp in length
* 11069 11168: gap of unknown length
* 11169 12400: contig of 1232 bp in length
* 12401 12500: gap of unknown length
* 12501 13625: contig of 1125 bp in length
* 13626 13725: gap of unknown length
* 13726 15211: contig of 1486 bp in length
* 15212 15311: gap of unknown length
* 15312 16498: contig of 1187 bp in length
* 16499 16598: gap of unknown length
* 16599 17661: contig of 1063 bp in length
* 17662 17761: gap of unknown length
* 17762 19412: contig of 1651 bp in length
* 19413 19512: gap of unknown length
* 19513 20847: contig of 1335 bp in length
* 20848 20948: gap of unknown length
* 20949 22030: contig of 1083 bp in length
* 22031 22130: gap of unknown length
* 22131 23435: contig of 1305 bp in length
* 23436 23535: gap of unknown length
* 23536 24609: contig of 1074 bp in length
* 24610 24709: gap of unknown length
* 24710 26236: contig of 1527 bp in length.

```

```

FEATURES
  source
    1. .26236
    /organism="Rattus norvegicus"
    /db_xref="taxon:10116"
    /clone="CH230-18605"
BASE COUNT  7879 a 4887 c 4901 g 6949 t 1620 others
ORIGIN

```

```

Query Match          93.8%: Score 15; DB 2; Length 26236;
Best Local Similarity 100.0%: Pred. No. 4.3e+02;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

Oy 2 cttcaccaccccc 16
    |||||||||
Db 6101 CTTTCACCAACCCCC 6115

```

```

RESULT 11
AC005942 26458 bp DNA linear PRI 06-JUN-2000
LOCUS Homo sapiens chromosome Y, clone hCIT.298_B_15, complete sequence.
AC005942
AC005942.2 GI:4263846
VERSION HTG.
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens

```

```

REFERENCE 1
AUTHORS Mammalia; Eutheria; Chordata; Craniata; Vertebrata; Euteleostomi;
1 (bases 1 to 26458)
Birren,B., Linton,L., Nusbaum,C., Page,D. and Lander,E.
Homo sapiens chromosome Y, clone hCIT.298_B_15
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 26458)
Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,M.,
Baker,J., Baldwin,J., Barna,N., Beckertly,R., Benn,J., Boutwell,C.,
Brown,A., Castle,A., Cerny,J., Cerny,N., Colangelo,M., Collins,S.,

```

```

TITLE
JOURNAL
REFERENCE
AUTHORS
Submitted (06-NOV-1998) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
3 (bases 1 to 26458)
Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,M.,
Baker,J., Baldwin,J., Barna,N., Beckertly,R., Benn,J., Brown,A.,
Castle,A., Cerny,J., Colangelo,M., Collins,S., Collamore,A.,
Cooke,P., Dearrellano,K., Depayre,E., Devon,K., Dewar,K.,
Donelan,L., Doyle,M., Ferreira,P., Fitzhugh,W., Forrest,C.,
Funke,R., Gage,D., Galagan,J., Gardyna,S., Gilbert,D., Grant,G.,
Hagos,B., Heaford,A., Horton,L., Howland,J.C., Jones,C., Kann,L.,
Karatas,A., Lehoczy,J., Lieu,C., Locke,K., Macdonald,P.,
Marquis,N., McEwan,P., McGurk,A., McKernan,K., McLaughlin,J.,
Meldrim,J., Molla,M., Morris,W., Morrow,J., Mychaleckyj,J.,
Naylor,J., Nilloff,M., O'Connor,T., O'Donnell,P., Pavlin,B.,
Peterson,K., Pollara,V., Riley,R., Roberts,D., Roy,A., Severy,P.,
Stange-Thomann,N., Stojanovic,N., Stone,C., Subramanian,A.,
Testaye,S., Tortuella-Miller,I., Vassiliev,H., Vo,A., Wagner,A.,
Wheeler,J., Wu,X., Wyman,D., Ye,W.J. and Zody,M.
Direct Submission
Submitted (24-FEB-1999) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
4 (bases 1 to 26458)
Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,M.,
Baker,J., Baldwin,J., Barna,N., Beckertly,R., Benn,J., Brown,A.,
Castle,A., Cerny,J., Colangelo,M., Collins,S., Collamore,A.,
Cooke,P., Dearrellano,K., Depayre,E., Devon,K., Dewar,K.,
Donelan,L., Doyle,M., Ferreira,P., Fitzhugh,W., Forrest,C.,
Funke,R., Gage,D., Galagan,J., Gardyna,S., Gilbert,D., Grant,G.,
Hagos,B., Heaford,A., Horton,L., Howland,J.C., Jones,C., Kann,L.,
Karatas,A., Lehoczy,J., Lieu,C., Locke,K., Macdonald,P.,
Marquis,N., McEwan,P., McGurk,A., McKernan,K., McLaughlin,J.,
Meldrim,J., Molla,M., Morris,W., Morrow,J., Mychaleckyj,J.,
Naylor,J., Nilloff,M., O'Connor,T., O'Donnell,P., Pavlin,B.,
Peterson,K., Pollara,V., Riley,R., Roberts,D., Roy,A., Severy,P.,
Stange-Thomann,N., Stojanovic,N., Stone,C., Subramanian,A.,
Testaye,S., Tortuella-Miller,I., Vassiliev,H., Vo,A., Wagner,A.,
Wheeler,J., Wu,X., Wyman,D., Ye,W.J. and Zody,M.
Direct Submission
Submitted (13-OCT-1999) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
5 (bases 1 to 26458)
Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,M.,
Baker,J., Baldwin,J., Barna,N., Beckertly,R., Benn,J., Brown,A.,
Castle,A., Cerny,J., Colangelo,M., Collins,S., Collamore,A.,
Cooke,P., Dearrellano,K., Depayre,E., Devon,K., Dewar,K.,
Donelan,L., Doyle,M., Ferreira,P., Fitzhugh,W., Forrest,C.,
Funke,R., Gage,D., Galagan,J., Gardyna,S., Gilbert,D., Grant,G.,
Hagos,B., Heaford,A., Horton,L., Howland,J.C., Jones,C., Kann,L.,
Karatas,A., Lehoczy,J., Lieu,C., Locke,K., Macdonald,P.,
Marquis,N., McEwan,P., McGurk,A., McKernan,K., McLaughlin,J.,
Meldrim,J., Molla,M., Morris,W., Morrow,J., Mychaleckyj,J.,
Naylor,J., Nilloff,M., O'Connor,T., O'Donnell,P., Pavlin,B.,
Peterson,K., Pollara,V., Riley,R., Roberts,D., Roy,A., Severy,P.,
Stange-Thomann,N., Stojanovic,N., Stone,C., Subramanian,A.,
Testaye,S., Tortuella-Miller,I., Vassiliev,H., Vo,A., Wagner,A.,
Wheeler,J., Wu,X., Wyman,D., Ye,W.J. and Zody,M.
Direct Submission
Submitted (06-JUN-2000) Whitehead Institute/MIT Center for Genome

```

```

TITLE
JOURNAL
Submitted (06-JUN-2000) Whitehead Institute/MIT Center for Genome

```

COMMENT Research, 320 Charles Street, Cambridge, MA 02141, USA  
On Feb 24, 1999 this sequence version replaced gi:4225939.  
All repeats were identified using RepeatMasker: Smit, A.F.A. &  
Green, P. (1996-1997)  
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>.

FEATURES Location/Qualifiers

Source  
1..26458  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="Y"  
/map="Y"  
/clone="hCITF.298.B.15"  
/clone\_lib="CITFC Human BAC"  
/rpt\_family="AluY"  
complement(1..123)  
/rpt\_family="AluY"  
complement(161..437)  
/rpt\_family="L1R9"  
complement(415..493)  
/rpt\_family="L1R9"  
complement(494..736)  
/rpt\_family="AluSp"  
complement(737..791)  
/rpt\_family="L1R9"  
complement(780..821)  
/note="Single-stranded coverage."  
complement(792..1106)  
/rpt\_family="AluSg"  
complement(1107..1392)  
/rpt\_family="L1R9"  
1396..1784  
/rpt\_family="MER4B-internal"  
1785..2001  
/rpt\_family="MER30"  
2002..2098  
/rpt\_family="MER4B-internal"  
complement(2099..2400)  
/rpt\_family="AluSx"  
2401..2732  
/rpt\_family="MER4B-internal"  
complement(2734..2921)  
/rpt\_family="THE1B"  
2898..3456  
/rpt\_family="MER4A"  
complement(3480..3538)  
/rpt\_family="MER72"  
3539..3702  
/rpt\_family="MER4B-internal"  
complement(3715..3974)  
/rpt\_family="MER4A"  
complement(4045..4270)  
/rpt\_family="MER4A"  
4271..4297  
/rpt\_family="TTTG)n"  
complement(4298..4393)  
/rpt\_family="MER4A"  
4413..4506  
/rpt\_family="MER4B-internal"  
complement(4523..4738)  
/rpt\_family="AluSg"  
4739..4907  
/rpt\_family="MER4B-internal"  
complement(4908..4926)  
/rpt\_family="MER50"  
4927..5229  
/rpt\_family="AluSg"  
complement(5230..5789)  
/rpt\_family="MER50"  
5792..6033  
/rpt\_family="MER4D"  
6085..6454  
/rpt\_family="MER4B-internal"  
6455..6625  
/rpt\_family="MER4-internal"

repeat\_region 6631..6844  
/rpt\_family="MER4-internal"  
complement(6845..7155)  
/rpt\_family="AluSg"  
7156..7355  
/rpt\_family="MER4-internal"  
7475..7577  
/rpt\_family="MER4-internal"  
7798..7890  
/rpt\_family="TTA)n"  
complement(8256..8811)  
/rpt\_family="MER61E"  
9152..9425  
/rpt\_family="AluJo"  
9509..9568  
/rpt\_family="AT-rich"  
9848..10031  
/rpt\_family="TTATATG)n"  
10219..10440  
/rpt\_family="TTATATG)n"  
complement(10441..10533)  
/rpt\_family="AluSg/x"  
11092..11375  
/rpt\_family="AluSg"  
11376..11398  
/rpt\_family="CAA)n"  
11425..11451  
/rpt\_family="AT-rich"  
11961..12074  
/rpt\_family="AluJo/FRAM"  
12180..12206  
/rpt\_family="AT-rich"  
12320..12403  
/rpt\_family="TTATG)n"  
complement(12784..12954)  
/rpt\_family="AluSg/x"  
complement(13050..13226)  
repeat\_region

Query Match 93.8%; Score 15; DB 9; Length 26458;  
Best Local Similarity 100.0%; Pred. No. 4.3e+02;  
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ccttcaccacccc 15  
Db 21811 CCTTACCAACCCC 21825

RESULT 12  
HS212J10  
LOCUS HS212J10 30208 bp DNA linear PRI 23-NOV-1999  
DEFINITION Human DNA sequence from clone 212J10 on chromosome Xq25-26.3.  
ACCESSION AL034401  
VERSION AL034401.1 GI:4071037  
KEYWORDS HTG.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
REFERENCE 1 (bases 1 to 30208)  
AUTHORS Grafham, D.  
TITLE Direct Submission  
JOURNAL Submitted (05-JAN-1999) Sanger Centre, Hinxton, Cambridgeshire,  
CB10 1SA, UK. E-mail enquiries: humquerry@sanger.ac.uk  
Requests: clonerequest@sanger.ac.uk  
On Dec 29, 1998 this sequence version replaced gi:4007136.  
During sequence assembly data is compared from overlapping clones.  
When differences are found these are annotated as variations  
together with a note of the overlapping clone name. Note that the  
variation annotation may not be found in the sequence submission  
corresponding to the overlapping clone, as we submit sequences with  
only a small overlap as described above.  
IMPORTANT: This sequence is not the entire insert of clone 212J10.

COMMENT

It may be shorter because we only sequence overlapping sections once, or longer because we arrange for a small overlap between neighbouring submissions. The true left end of clone 448E20 (297196) is at 30109 in this sequence. The true right end of clone 417G15 (AL009174) is at 100 in this sequence.

This sequence has been finished according to sequence map criteria as follows. An attempt is made to resolve all sequencing problems, such as compressions and repeats, but not necessarily within known annotated human repeat sequence elements (e.g. Alu). Where the sequence is ambiguous, there is an annotation using the 'unsure' feature key.

This sequence was generated from part of bacterial clone contigs of human chromosome X, constructed by the Sanger Centre Chromosome X Mapping Group. Further information can be found at

<http://www.sanger.ac.uk/HGP/ChrX>  
212J10 is from the library RPEC1 constructed at the Roswell Park Cancer Institute by the group of Pieter de Jong. For further details see <http://bacpac.med.buffalo.edu/VECTOR.pcyPAC2>.

## FEATURES

## SOURCE

```
1. 30208
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="X"
/map="q25-26.3"
/clone="RP1-212J10"
/clone_1lb="RRC1-1"
1166..1275
repeat_region
/feature="L2 repeat: matches 2636..2738 of consensus"
1307..1423
repeat_region
/feature="MIR repeat: matches 19..134 of consensus"
1586..1620
repeat_region
/feature="L2 repeat: matches 2515..2551 of consensus"
1621..1897
repeat_region
/feature="AluX repeat: matches 32..312 of consensus"
1898..2108
repeat_region
/feature="L2 repeat: matches 2551..2745 of consensus"
complement(2474..2892)
/feature="match: ESTs R94487 A1022438"
3566..4006
misc_feature
/feature="match: GSS B95157"
4688..4806
repeat_region
/feature="L2 repeat: matches 2575..2698 of consensus"
5684..5993
repeat_region
/feature="AluY repeat: matches 2..311 of consensus"
6908..7088
repeat_region
/feature="L2 repeat: matches 2495..2679 of consensus"
7298..7608
repeat_region
/feature="AluX repeat: matches 1..309 of consensus"
8688..8859
repeat_region
/feature="MER45 repeat: matches 1..173 of consensus"
8860..9149
repeat_region
/feature="AluB repeat: matches 1..290 of consensus"
9245..9323
repeat_region
/feature="L2 repeat: matches 2584..2676 of consensus"
12097..12126
repeat_region
/feature="10 copies 3 mer aac 100% conserved"
12662..12972
repeat_region
/feature="AluSq repeat: matches 3..309 of consensus"
13719..13770
repeat_region
/feature="26 copies 2 mer ca 87% conserved"
14015..14135
repeat_region
/feature="MER5B repeat: matches 12..129 of consensus"
14161..14346
repeat_region
/feature="L2 repeat: matches 1629..1839 of consensus"
15319..15553
repeat_region
/feature="MIR repeat: matches 20..261 of consensus"
16380..16697
repeat_region
/feature="AluSq repeat: matches 1..312 of consensus"
16805..17219
repeat_region
/feature="WSTB repeat: matches 1..426 of consensus"
17435..17491
repeat_region
/feature="MIR repeat: matches 201..255 of consensus"
```

```
repeat_region 17550..17735
/feature="MER5A repeat: matches 3..189 of consensus"
18324..18488
repeat_region
/feature="MIR repeat: matches 2..218 of consensus"
19212..19275
repeat_region
/feature="2 copies 32 mer 92% conserved"
20376..20588
repeat_region
/feature="AluY repeat: matches 3..215 of consensus"
20589..20630
repeat_region
/feature="Alu repeat: matches 268..307 of consensus"
20774..20914
repeat_region
/feature="MER94 repeat: matches 1..134 of consensus"
21170..21251
repeat_region
/feature="41 copies 2 mer tt 76% conserved"
24307..24617
repeat_region
/feature="AluX repeat: matches 1..311 of consensus"
25794..26104
repeat_region
/feature="AluB repeat: matches 1..307 of consensus"
26242..26339
repeat_region
/feature="MIR repeat: matches 164..261 of consensus"
26347..26505
repeat_region
/feature="L2 repeat: matches 1450..1604 of consensus"
28752..29001
repeat_region
/feature="L1PA3 repeat: matches 5894..6146 of consensus"
29377..29494
repeat_region
/feature="MIR repeat: matches 8..124 of consensus"
29510..29698
repeat_region
/feature="MIR repeat: matches 12..212 of consensus"
BASE COUNT 9229 a 6309 c 5771 g 8899 t
ORIGIN
```

Query Match 93.8% Score 15: DB 9: Length 30208;  
Best Local Similarity 100.0%: Pred. No. 4.3e+02;  
Matches 15: Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
QY 1 ccttcaccacccc 15
|||||
DB 1968 CCTTCCACCAACCC 1982
```

```
RESULT 13
AC099662 46363 bp DNA linear HTG 17-NOV-2001
LOCUS Rattus norvegicus clone CH230-1D24, *** SEQUENCING IN PROGRESS ***
DEFINITION 40 unordered pieces.
ACCESSION AC099662
VERSION AC099662.1 GI:16973712
KEYWORDS HTG; HTGS; PHASE1.
SOURCE Norway rat.
ORGANISM Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
```

REFERENCE  
AUTHORS

1 (bases 1 to 46363)  
Muzny,D.M., Adams,C., Adio-Oduola,B., Alf-ossman,F.R., Allen,C.,  
Albrooks,S.L., Amaratunge,H.C., Are,J.R., Banks,T., Barbara,J.,  
Benton,J., Bimaga,K., Blankenburg,K., Bonnin,D., Bouck,J.,  
Bowle,S., Bivenga,M., Brown,E., Brown,M., Bryant,N.P., Bunay,C.,  
Burkett,C., Burrell,K.L., Byrd,N.C., Carron,T.F.,  
Carter,M., Cavazos,S.R., Chacko,J., Chavez,D., Chen,G., Chen,R.,  
Chen,Z., Chowdhry,I., Christopoulos,C., Cleveland,C.D., Cox,C.,  
Coyle,M.D., Dathorne,S.R., David,R., Davila,M.L., Davis,C.,  
Davy-Carroll,L., Dederich,D.A., Delaney,K.R., Delgado,O.,  
Denn,A.L., Ding,Y., Dinh,H.H., Douthwaite,K.J., Draper,H.,  
Dugan-Rocha,S., Durbin,K.J., Earnhart,C., Edgar,D., Edwards,C.C.,  
Elhaj,C., Escotlo,M., Falls,T., Ferriguto,D., Flagg,N., Ford,J.,  
Foster,P., Frantz,P., Gabisi,A., Gao,J., Garcia,A., Garner,T.,  
Garza,N., Gill,R., Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S.,  
Hamilton,K., Harris,C., Harris,K., Hart,M., Haylak,P., Hawes,A.,  
Hernandez,J., Hernandez,O., Hodgson,A., Hughes,M., Holloway,C.,  
Hollins,B., Homs,F., Howard,S., Huber,J., Hulyk,S., Hume,J.,  
Jackson,L.E., Jacobson,B., Jia,Y., Johnson,R., Jolivet,S.,

Joudah, S., Karlsson, E., Kelly, S., Khan, U., King, L., Korvah, J., Kovar, C., Kravic, D., Kuresh, A., Landry, N., Leal, B., Lewis, L. C., Lewis, L., Li, J., Li, Z., Lichter, O., Lieu, C., Liu, J., Liu, W., Louised, H., Lozano, R. J., Lu, X., Lucier, A., Lucier, R., Luna, R., Ma, J., Maheshwari, M., Mapua, P., Martin, R., Martindale, A., Martinez, E., Massey, E., Mawhney, E., McLeod, M. P., Meador, M., Mel, G., Metzger, M., Miner, G., Miner, Z., Mitchell, T., Mohabbat, K., Morgan, M., Morris, S., Moser, M., Neal, D., Newton, J., Newton, N., Nguyen, A., Nguyen, N., Nickerson, E., Nwokweto, S., Ogutu, M., Okunolu, G., Oragunye, N., Oviedo, R., Pace, A., Payton, B., Peery, J., Perez, L., Peters, L., Pickens, R., Primus, E., Pu, L. L., Quiles, M., Ren, Y., Rives, M., Rojas, A., Rojupokan, I., Rolfe, M., Ruiz, S., Savary, G., Scherer, S., Scott, G., Shen, H., Shooshari, N., Sisson, I., Sodergren, E., Sonalke, T., Sparks, A., Stanley, H., Stone, I., Sutton, A., Svatek, A., Taber, P., Tamerisa, A., Tamerisa, K., Tang, H., Tansey, J., Taylor, C., Taylor, T., Telford, B., Thomas, N., Thomas, S., Usmani, K., Vasquez, L., Vera, V., Villalob, D., Vinson, R., Wall, R., Wang, S., Ward-Moore, S., Warren, R., Washington, C., Watlington, S., Williams, G., Williamson, A., Wleczek, R., Wooden, S., Worley, K., Wu, C., Wu, Y., Wu, Y. F., Zhou, J., Zorrilla, S., Nelson, D., Weinstein, G., and Gibbs, R.

Direct Submission  
Unpublished  
2 (bases 1 to 46363)  
Worley, K.C.  
Submitted (17-NOV-2001) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA

----- Genome Center of Medicine  
Center: Baylor College of Medicine  
Center code: BCM  
Web site: <http://www.hgsc.bcm.tmc.edu/>  
Contact: hgsc-help@bcm.tmc.edu

----- Project Information  
Center project name: TUPM  
Center clone name: CH230-ID24

----- Summary Statistics  
Sequencing vector: Plasmid; M77789  
Chemistry: Dye-terminator Big Dye; 99% of reads  
Assembly program: Phrap; version 0.990329  
Consensus quality: 19414 bases at least Q40  
Consensus quality: 22250 bases at least Q30  
Consensus quality: 24040 bases at least Q20  
Estimated insert size: 7952; sum-of-contigs estimation  
Quality coverage: 0x in Q20 bases; agarose-fp estimation  
Quality coverage: 0.1x in Q20 bases; sum-of-contigs estimation

-----  
\* NOTE: Estimated insert size may differ from sequence length  
\* (see [http://www.hgsc.bcm.tmc.edu/docs/Genbank\\_draft\\_data.html](http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html)).  
\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 40 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will  
\* be preserved.

1  
898 997: contig of 897 bp in length  
998 997: gap of unknown length  
1916 1915: contig of 918 bp in length  
2016 2015: gap of unknown length  
2962 2961: contig of 946 bp in length  
3062 3061: gap of unknown length  
3971 3970: contig of 909 bp in length  
4071 4070: gap of unknown length  
4961 4960: contig of 890 bp in length  
5061 5060: gap of unknown length  
5990 5989: contig of 929 bp in length  
6090 6089: gap of unknown length  
7034 7033: contig of 944 bp in length  
7134 7133: gap of unknown length  
8079: contig of 946 bp in length

8080 8179: gap of unknown length  
8180 9076: contig of 897 bp in length  
9077 9176: gap of unknown length  
9177 10077: contig of 901 bp in length  
10078 10177: gap of unknown length  
10178 11122: contig of 945 bp in length  
11123 11222: gap of unknown length  
11223 12158: contig of 936 bp in length  
12159 12259: gap of unknown length  
12259 13180: contig of 922 bp in length  
13181 13280: gap of unknown length  
13281 14190: contig of 910 bp in length  
14191 14290: gap of unknown length  
14291 16032: contig of 1742 bp in length  
16033 16132: gap of unknown length  
16133 17561: contig of 1429 bp in length  
17562 17661: gap of unknown length  
17662 17927: contig of 266 bp in length  
17928 18027: gap of unknown length  
18028 19429: contig of 1402 bp in length  
19430 19529: gap of unknown length  
19530 20481: contig of 952 bp in length  
20482 20581: gap of unknown length  
20582 21432: contig of 851 bp in length  
21433 21532: gap of unknown length  
21533 22815: contig of 1283 bp in length  
22816 22915: gap of unknown length  
22916 24328: contig of 1313 bp in length  
24329 25333: gap of unknown length  
25334 25433: gap of unknown length  
25434 26350: contig of 917 bp in length  
26351 26450: gap of unknown length  
26451 27341: contig of 891 bp in length  
27342 27441: gap of unknown length  
27442 28606: contig of 1165 bp in length  
28607 28706: gap of unknown length  
28707 29621: contig of 915 bp in length  
29622 29721: gap of unknown length  
29722 30581: contig of 860 bp in length  
30582 30681: gap of unknown length  
30682 32084: contig of 1403 bp in length  
32085 32184: gap of unknown length  
32185 33096: contig of 912 bp in length  
33097 33196: gap of unknown length  
33197 34478: contig of 1282 bp in length  
34479 34578: gap of unknown length  
34579 36167: contig of 1559 bp in length  
36168 36267: gap of unknown length  
36268 37380: contig of 1113 bp in length  
37381 37480: gap of unknown length  
37481 38539: contig of 1059 bp in length  
38540 38639: gap of unknown length  
38640 40235: contig of 1556 bp in length  
40236 40335: gap of unknown length  
40336 41516: contig of 1181 bp in length  
41517 42512: contig of 896 bp in length  
42513 42612: gap of unknown length  
42613 43771: contig of 1159 bp in length  
43772 43871: gap of unknown length  
43872 45209: contig of 1338 bp in length  
45210 45309: gap of unknown length  
45310 46363: contig of 1054 bp in length.

Location/Qualifiers  
1. 46363  
/organism="Rattus norvegicus"  
/db\_xref="taxon:10116"  
/clone="CH230-ID24"

BASE COUNT 11211 a 11090 c 8068 g 11580 t 4414 others  
ORIGIN





```

* 29428 29327: gap of 100 bp
* 29328 30008: contig of 681 bp in length
* 30009 30108: gap of 100 bp
* 30109 30769: contig of 661 bp in length
* 30770 30869: gap of 100 bp
* 30870 31564: contig of 695 bp in length
* 31565 31664: gap of 100 bp
* 31665 32350: contig of 686 bp in length
* 32351 32450: gap of 100 bp
* 32451 33143: contig of 693 bp in length
* 33144 33243: gap of 100 bp
* 33244 33953: contig of 710 bp in length
* 33954 34053: gap of 100 bp
* 34054 34750: contig of 697 bp in length
* 34751 34850: gap of 100 bp
* 34851 35562: contig of 712 bp in length
* 35563 35662: gap of 100 bp
* 35663 36361: contig of 699 bp in length
* 36362 36461: gap of 100 bp
* 36462 37150: contig of 689 bp in length
* 37151 37250: gap of 100 bp
* 37251 37931: contig of 681 bp in length
* 37932 38031: gap of 100 bp
* 38032 38722: contig of 691 bp in length
* 38723 38822: gap of 100 bp
* 38823 39522: contig of 700 bp in length
* 39523 40322: contig of 700 bp in length
* 40323 40422: gap of 100 bp
* 40423 41118: contig of 696 bp in length
* 41119 41218: gap of 100 bp
* 41219 41924: contig of 706 bp in length
* 41925 42024: gap of 100 bp
* 42025 42710: contig of 686 bp in length
* 42711 42810: gap of 100 bp
* 42811 43503: contig of 693 bp in length
* 43504 43603: gap of 100 bp
* 43604 44289: contig of 686 bp in length
* 44290 44389: gap of 100 bp
* 44390 45073: contig of 684 bp in length
* 45074 45173: gap of 100 bp
* 45174 45782: contig of 609 bp in length
* 45783 45882: gap of 100 bp
* 45883 46576: contig of 694 bp in length
* 46577 46676: gap of 100 bp
* 46678 47363: contig of 687 bp in length
* 47364 47463: gap of 100 bp
* 47464 48161: contig of 698 bp in length
* 48162 48261: gap of 100 bp
* 48262 48976: contig of 715 bp in length
* 48977 49076: gap of 100 bp
* 49078 49778: contig of 702 bp in length
* 49779 49878: gap of 100 bp
* 49879 50597: contig of 719 bp in length
* 50598 50697: gap of 100 bp
* 50698 51414: contig of 717 bp in length
* 51415 51514: gap of 100 bp
* 51515 52214: contig of 700 bp in length
* 52215 52314: gap of 100 bp
* 52315 53010: contig of 696 bp in length
* 53011 53110: gap of 100 bp
* 53111 53796: contig of 686 bp in length
* 53797 53896: gap of 100 bp
* 53897 54584: contig of 688 bp in length
* 54585 54684: gap of 100 bp

```

Query Match 93.8%, Score 15, DB 2, Length 62490;  
 Best Local Similarity 100.0%, Pred. No. 4.4e+02;

Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

0y 1 ccttcacacccc 15  
 Db 48418 CCTTCACCAACCCC 48432

```

RESULT 15
AC098003
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
1 (bases 1 to 76122)
Munzy,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C.,
Alsbrooks,S.L., Amaratunga,H.C., Are,J.R., Banks,T., Barbara,J.,
Benton,J., Blinag,K., Blankenburg,K., Bonnin,D., Bouck,J.,
Bowle,S., Brileva,M., Brown,E., Brown,M., Bryant,N.P., Buhay,C.,
Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C., Carron,T.F.,
Carter,M., Cavazos,S.R., Chacko,J., Chavez,D., Chen,G., Chen,R.,
Chen,Z., Chowdhry,I., Christopoulos,C., Cleveland,C.D., Cox,C.,
Coyle,M.D., Dethorne,S.R., David,R., Davila,M.L., Davis,C.,
Davy-Carroll,U., Dederich,D.A., Delaney,K.R., Delgado,O.,
Denn,A.L., Ding,Y., Dinh,H.H., Doutheaite,K.J., Draper,H.,
Dugan-Rocha,S., Durbin,K.J., Barnhart,C., Edgar,D., Edwards,C.C.,
Elhaj,C., Escotto,M., Falls,T., Ferraguto,D., Flagg,N., Ford,J.,
Foster,P., Frantz,P., Gabisi,A., Gao,J., Garcia,A., Garner,T.,
Garza,N., Gill,R., Gorrell,J.H., Gueraa,M., Gunaratne,P., Hale,S.,
Hamilton,K., Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A.,
Hernandez,J., Hernandez,O., Hodgson,A., Hogues,M., Holloway,C.,
Hollins,B., Homs,F., Howard,S., Huber,D., Hulyk,S., Hume,J.,
Jackson,L.E., Jacobson,B., Jia,Y., Johnson,R., Joliver,S.,
Jondah,S., Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J.,
Kovar,C., Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,I.C.,
Lewis,L., Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W.,
Louiaged,H., Lozado,R.J., Lu,X., Lucier,A., Lucinda,R., Luna,R.,
Ma,J., Maheshwari,M., Mapa,P., Martin,R., Martindale,A.,
Martinez,E., Massey,E., Mawhney,E., Mcleod,M.P., Meador,M.,
Mel,G., Metzker,M., Miner,G., Miner,Z., Mitchell,T., Mohabbat,K.,
Morgan,M., Morris,S., Moser,M., Neal,D., Newton,J., Newton,N.,
Nguyen,A., Nguyen,N., Nguyen,N., Nickerson,E., Nwokemwo,S.,
Ogun,M., Okunonu,G., Orangunye,N., Oviedo,R., Pace,A., Payton,B.,
Peery,J., Perez,L., Peters,L., Pickens,R., Primus,E., Pu,L.L.,
Quiles,M., Ren,Y., Rives,M., Rojas,A., Rojubenok,I., Rolfe,M.,
Ruiz,S., Savery,G., Scherer,S., Scott,G., Shen,H., Shoostari,N.,
Sisson,I., Sodergren,E., Sonalke,T., Sparks,A., Stanley,H.,
Stone,H., Sutton,A., Svatek,A., Tabot,P., Tamerisa,A., Tamerisa,K.,
Tang,H., Tansey,J., Taylor,C., Taylor,T., Telford,B., Thomas,N.,
Thomas,S., Usmani,K., Vasquez,L., Vera,V., Villalón,D., Vinson,R.,
Wall,R., Wang,S., Ward-Moore,S., Warren,A., Wazhington,C.,
Watlington,S., Williams,G., Williamson,A., Wlezyk,R., Wooden,S.,
Worley,K., Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D.,
Weinstock,G. and Gibbs,R.
Direct Submission
Unpublished
2 (bases 1 to 76122)
Worley,K.C.
Direct Submission
Submitted (23-Oct-2001) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
On Dec 20, 2001 this sequence version replaced gi:17062261.
----- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: http://www.hgsc.bcm.tmc.edu/
Contact: hgsc-help@bcm.tmc.edu
----- Project Information
Center project name: GRWG
Center clone name: CH230-103F13
----- Summary Statistics

```



---



PR 12-NOV-1997; 97US-0969046.  
 XX (UNMT ) UNIV MICHIGAN TECHNOLOGICAL.  
 XX  
 PI Chiang VIC, Hu W, Tsai C;  
 DR WPI; 1999-327394/27.  
 XX  
 PT Altering properties of plants by modulating 4-coumarate co-enzyme A  
 PT ligase  
 XX  
 PS Disclosure; Page 71; 73pp; English.  
 XX  
 CC This is the nucleotide sequence of the promoter region of the  
 CC quaking aspen (*Populus tremuloides* Michx.) Pt4Cl1 gene (see  
 CC also AX58642) that codes for 4-coumarate coenzyme A ligase 4Cl1 (see  
 CC AY06092). The promoter DNA was isolated from an aspen genomic  
 CC library by screening with Pt4Cl1 cDNA. The promoter drives gene  
 CC expression exclusively in xylem tissue. It can be used to  
 CC manipulate gene expression, and hence to engineer traits of  
 CC interest, in the xylem tissue of target plants, e.g. to manipulate  
 CC lignin content or structure, or to enhance growth, cellulose  
 CC content or other value-added wood qualities. Plants with altered  
 CC contents of lignin and/or cellulose can be processed more  
 CC efficiently, e.g. for pulp production, with reduced costs and  
 CC pollution associated with lignin removal.  
 XX  
 SQ Sequence 1172 BP; 399 A; 225 C; 180 G; 368 T; 0 other;

Query Match 100.0%; Score 16; DB 20; Length 1172;  
 Best Local Similarity 100.0%; Pred. No. 64;  
 Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 ccttcacacacccc 16  
 |||||||  
 DB 1018 ccttcacacacccc 1033

## RESULT 2

AA50528  
 ID AA50528 standard; DNA; 1440 BP.

XX  
 AC AA50528;

XX  
 DT 09-DEC-1991 (first entry)

XX  
 DE Parasponia rhizobium ANU289 nifd gene.

XX  
 KW nifd gene; nitrogenase; enzyme; nitrogen-fixation;  
 KM strain improvement; plasmid pBR289nif-2; plasmid pR329nif-2; ss.

XX  
 OS Parasponia rhizobium.

XX  
 FH Key Location/Qualifiers  
 FT CDS 175..1434  
 FT /\*tag= a

XX  
 PN EPI30047-A.

XX  
 PD 02-JAN-1985.

XX  
 PF 21-JUN-1984; 84EP-0304191.

XX  
 PR 22-JUN-1983; 83US-0506676.

XX  
 PR 16-OCT-1987; 87US-0109868.

XX  
 PA (AGRI-) AGRIGENETICS RES AS.

XX  
 PI Shine J, Rolfe BG, Scott KF;

XX  
 DR WPI; 1985-007981/02.

XX  
 DR P-PDB; AAP50775.

XX  
 PT Bacterial strain contg. recombinant DNA fragment - esp. in  
 PT Rhizobium strains for improved nitrogen fixation  
 XX  
 PS Disclosure; Fig 5; 52pp; English.

XX  
 CC When this gene is expressed in a Rhizobium sp., the bacterium  
 CC produces useful products and plants may have their properties  
 CC improved, e.g. the rate, quality and efficiency of the nitrogen-  
 CC fixation process, especially in the root nodules of Rhizobium  
 CC strains. See also AA50526-7 and AAP50773-4.

XX  
 SQ Sequence 1440 BP; 355 A; 509 C; 283 G; 293 T; 0 other;

Query Match 93.8%; Score 15; DB 6; Length 1440;  
 Best Local Similarity 100.0%; Pred. No. 2; 1e+02;  
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2 cttcaccacacccc 16  
 |||||||  
 DB 294 cttcaccacacccc 308

## RESULT 3

AA513959/C  
 ID AA513959 standard; DNA; 3084 BP.

XX  
 AC AA513959;

XX  
 DT 31-MAR-1999 (first entry)

XX  
 DE H. pylori GHPO 741 gene.

XX  
 KW GHPO protein; Helicobacter infection; gastroduodenal disease; gastritis;  
 KM peptic ulcer disease; ss.

XX  
 OS Helicobacter pylori.

XX  
 FH Key Location/Qualifiers  
 FT CDS 49..3030  
 FT /\*tag= a

XX  
 PN WO9843478-A1.

XX  
 PD 08-OCT-1998.

XX  
 PF 01-APR-1998; 98WO-US06371.

XX  
 PR 29-JUL-1997; 97US-0902615.

XX  
 PR 01-APR-1997; 97US-0833457.

XX  
 PR 24-JUN-1997; 97US-0881227.

XX  
 PA (HUMA-) HUMAN GENOME SCI INC.

XX  
 PA (INMR ) MERIEUX ORAVAX PASTEUR MERIEUX SERUMS.

XX  
 PI Al-Garawi A, Kleanthous H, Miller C, Oomen RP, Tomb J;

XX  
 DR WPI; 1998-542293/46.

XX  
 DR P-PDB; AA598240.

XX  
 PT New isolated Helicobacter polynucleotides - used to develop products  
 PT for the diagnosis, prevention and treatment of Helicobacter  
 PT infections and gastrointestinal diseases

XX  
 PS Claim 1; Page 200-204; 2054pp; English.

XX  
 CC This sequence represents a polynucleotide of the invention. It was  
 CC isolated from Helicobacter pylori and encodes a H.pylori GHPO protein.  
 CC The polypeptides can be used for preventing or treating Helicobacter  
 CC infections, and gastroduodenal diseases associated with these  
 CC infections, including acute, chronic, and atrophic gastritis, and peptic  
 CC ulcer diseases, e.g. gastric and duodenal ulcers. They can also be used

CC for the production of antibodies. The products can also be used for  
CC detection and diagnosis.  
XX  
SQ Sequence 3084 BP; 960 A; 589 C; 721 G; 814 T; 0 other;

Query Match 93.8%; Score 15; DB 19; Length 3084;  
Best Local Similarity 100.0%; Pred. No. 2.2e+02;  
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2 cttcaccaccccc 16  
|||||  
DB 1711 CTTTCACCAACCCCC 1697

RESULT 4  
AAS53676/c  
ID AAS53676 standard; DNA; 8673 BP.  
XX  
AC AAS53676;

DT 13-FEB-2002 (first entry)

DE Helicobacter pylori DNA for cellular proliferation protein #130.

XX Antisense; ds; prokaryotic cellular proliferation gene;  
KW antibiotic; antibacterial; drug design.

OS Helicobacter pylori.

XX MO200170955-A2.

XX 27-SEP-2001.

PD 21-MAR-2001; 2001WO-US09180.

XX 21-MAR-2000; 2000US-191078P.

PR 23-MAY-2000; 2000US-206848P.

PR 26-MAY-2000; 2000US-207727P.

XX 23-OCT-2000; 2000US-242578P.

PR 27-NOV-2000; 2000US-253625P.

PR 22-DEC-2000; 2000US-257931P.

PR 16-FEB-2001; 2001US-269308P.

XX (ELIT-) ELITRA PHARM INC.

PI Haseelbeck R, Ohlsen KL, Zyskind JW, Wall D, Trawick JD, Carr GJ;  
PI Yamamoto RT, Xu HH;

XX WPI; 2001-611495/70.

DR P-PSDB; AAU35817.

XX New polynucleotides for the identification and development of  
PT antibiotics, comprise sequences of antisense nucleic acids -

XX Claim 27; Seq ID No 7313; 511pp; English.

CC The invention relates to antisense inhibitors of genes essential to  
CC prokaryotic cellular proliferation, their use in identifying the

CC genes, their use in the discovery of novel antibiotics, the essential  
CC genes themselves and the encoded proteins. The prokaryotes used are

CC Escherichia coli, Staphylococcus aureus, Salmonella typhi, Klebsiella  
CC pneumoniae, Pseudomonas aeruginosa and Enterococcus faecalis. The

CC invention is also useful for the identification of potential new targets  
CC for antibiotic development. The antisense nucleic acids can also be used

CC to identify proteins used in proliferation, to express these proteins,  
CC and to obtain antibodies capable of binding to the expressed proteins.

CC The proteins can be used to screen compounds in rational drug discovery  
CC programmes. The antisense nucleic acid sequence is also useful to screen  
CC for homologous nucleic acids which are required for cell proliferation in  
CC a wide variety of organisms. The present sequence encodes an  
CC essential prokaryotic cellular proliferation protein.  
CC Note: The sequence data for this patent did not form part

CC of the printed specification, but was obtained in electronic  
CC format directly from WIPO at  
CC ftp.wipo.int/pub/published\_pct\_sequences.  
XX

SQ Sequence 8673 BP; 2673 A; 1595 C; 2181 G; 2224 T; 0 other;

Query Match 93.8%; Score 15; DB 23; Length 8673;  
Best Local Similarity 100.0%; Pred. No. 2.3e+02;  
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2 cttcaccaccccc 16  
|||||  
DB 1663 CTTTCACCAACCCCC 1649

RESULT 5  
AAI61371/c  
ID AAI61371 standard; DNA; 335913 BP.  
XX  
AC AAI61371;

XX 16-OCT-2001 (first entry)

DT Soybean 240017 region G3, SEQ ID NO: 2.

XX Soybean 240017 region G3, SEQ ID NO: 2.

XX Soybean: anthelmintic; gene therapy; soybean cyst nematode; SCN;

XX SCN resistance; rhg1; Rhg4; SCN resistant allele; plant breeding;

XX 240017 region G3; 318013 region A5; 515002 region G2; ds.

XX Glycine max.

XX MO200151627-A2.

XX 19-JUL-2001.

PD 05-JAN-2001; 2001WO-US00552.

XX 07-JAN-2000; 2000US-0174880.

XX (MONS) MONSANTO CO.

XX Hauge BM, Wang ML, Parsons JD, Parnell LD;

XX WPI; 2001-425872/45.

DR P-PSDB; AAM42214.

XX New purified nucleic acid for producing a soybean plant having soybean  
PT cyst nematode resistance and for use in plant breeding programs -

XX Claim 2; Page 204-400; 1353pp; English.

CC The invention relates to nucleic acid molecules from regions of the  
CC soybean genome which are associated with soybean cyst nematode (SCN)

CC resistance. The nucleic acids are used to transform plants, and can  
CC produce soybean plants having an rhg1 or an Rhg4 SCN resistant allele.

CC The nucleic acids can be used for investigating rhg1 or Rhg4 haplotypes  
CC of soybean plants and for introgressing SCN resistance or partial SCN

CC resistance into soybean plants. They can also be used in plant breeding  
CC programmes. The invention also relates to proteins encoded by such

CC nucleic acid molecules, as well as antibodies capable of recognizing  
CC these proteins. The present sequence is a nucleic acid molecule  
CC provided in the specification.

XX Sequence 335913 BP; 114579 A; 53403 C; 53026 G; 114905 T; 0 other;

Query Match 93.8%; Score 15; DB 22; Length 335913;  
Best Local Similarity 100.0%; Pred. No. 2.7e+02;  
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 ccttcaccaccccc 15  
|||||

DB 291134 CCTTCACCAACCC 291120

# RESULT 6

AA161372/c  
ID AA161372 standard; DNA: 335913 BP.

XX  
AC AA161372;

XX  
DT 16-OCT-2001 (first entry)

XX  
DE Soybean 240017 region G3, SEQ ID NO: 3.

XX  
KW Soybean; antihelminthic; gene therapy; soybean cyst nematode; SCN;  
KW resistance; rhg1; Rhg4; SCN resistant allele; plant breeding;  
KW 240017 region G3; 318013 region A3; 515002 region G2; ds.

XX  
OS Glycine max.

XX  
PN WO200151627-A2;

XX  
PD 19-JUL-2001.

XX  
PF 05-JAN-2001; 2001WO-US00552.

XX  
PR 07-JAN-2000; 2000US-0174880.

XX  
PA (MONS ) MONSANTO CO.

XX  
PI Hauge BM, Wang ML, Parsons JD, Parnell LD;

XX  
DR WPI; 2001-425872/45.

XX  
P-PSDB; AAM42215.

PT  
PT New purified nucleic acid for producing a soybean plant having soybean  
cyst nematode resistance and for use in plant breeding programs -

PS  
PS Claim 2; Page 400-595; 1353pp; English.

XX  
CC The invention relates to nucleic acid molecules from regions of the  
CC soybean genome, which are associated with soybean cyst nematode (SCN)  
CC resistance. The nucleic acids are used to transform plants, and can  
CC produce soybean plants having an rhg1 or an Rhg4 SCN resistant allele.  
CC The nucleic acids can be used for investigating rhg1 or Rhg4 haplotypes  
CC of soybean plants and for introgressing SCN resistance or partial SCN  
CC resistance into soybean plants. They can also be used in plant breeding  
CC programmes. The invention also relates to proteins encoded by such  
CC nucleic acid molecules, as well as antibodies capable of recognising  
CC these proteins. The present sequence is a nucleic acid molecule  
CC provided in the specification.

SQ Sequence 335913 BP; 114582 A; 53398 C; 53027 G; 114906 T; 0 other;

## Query Match

Best Local Similarity 93.8%; Score 15; DB 22; Length 335913;  
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 ccttcaccacccc 15  
|||||

DB 291134 CCTTCACCAACCC 291120

## RESULT 7

AA005460/c  
ID AA005460 standard; CDNA: 347 BP.

XX  
AC AA005460;

XX  
DT 17-JUL-2001 (first entry)

XX  
DE Human secreted protein-encoding gene 22 CDNA clone HT4ES80, SEQ ID NO:82.

KW Human; secreted protein; proliferative disorder; cancer; tumour;  
KW foetal abnormality; developmental abnormality; haematopoietic disorder;  
KW immune system disorder; AIDS; autoimmune disease; rheumatoid arthritis;  
KW inflammation; allergy; neurological disorder; Alzheimer's disease;  
KW Parkinson's disease; cognitive disorder; schizophrenia; asthma;  
KW skin disorder; psoriasis; sepsis; diabetes; atherosclerosis;  
KW cardiovascular disorder; angioecnic disorder; kidney disorder;  
KW gastrointestinal disorder; pregnancy-related disorder;  
KW endocrine disorder; infection; wound healing; vulnery;  
KW cell culture; chemotaxis; food additive; gene therapy;  
KW binding partner identification; chromosome 16; ss.

XX  
OS Homo sapiens.

XX  
FH Key Location/Qualifiers

FT CDS 1..213

FT /tag= a

FT /product= "Human secreted protein"

FT /transl\_except= (pos:58..60, aa:Xaa)  
FT /transl\_except= (pos:178..180, aa:Xaa)  
FT /note= "Xaa corresponds to any of the naturally occurring  
L-amino acids; CDS does not include start codon"

FT sig\_peptide 1..3

FT /tag= b

FT mat\_peptide 4..210

FT /tag= c

FT /product= "Mature human secreted protein"

XX  
PN WO200134623-A1.

XX  
PD 17-MAY-2001.

XX  
PF 01-NOV-2000; 2000WO-US30037.

XX  
PR 05-NOV-1999; 99US-0163577.

XX  
PA 30-JUN-2000; 2000US-0215137.

XX  
PA (HUMA-) HUMAN GENOME SCI INC.

XX  
PI Ruben SM, Komatsoulis GA, Moore PA;

XX  
DR WPI; 2001-316490/33.

XX  
P-PSDB; AAE01617.

PT  
PT Nucleic acids encoding 29 human secreted polypeptides, useful for  
preventing, diagnosing and/or treating e.g. cancers, Parkinson's  
disease and diabetic retinopathy -

PS  
PS Claim 1; Page 468; 535pp; English.

XX  
CC AA005389-AA005473 represent cDNAs corresponding to 29 human secreted  
CC protein genes, and AAE01546-AAE01630 represent the proteins they encode.  
CC AA001631-AAE01660 represent human secreted protein fragments or variants.  
CC The secreted proteins and their genes are useful for preventing,  
CC treating or ameliorating medical conditions, e.g., by protein or gene  
CC therapy. Pathological conditions can be diagnosed by determining the  
CC amount of the new protein in a sample or by determining the presence of  
CC mutations in the new genes. Specific uses are described for each of the  
CC 29 genes, based on the tissues in which they are most highly expressed,  
CC and include developing products for the diagnosis or treatment of  
CC proliferative disorders, cancer, tumours, foetal and developmental  
CC abnormalities, haematopoietic disorders, diseases of the immune system,  
CC AIDS, autoimmune diseases (e.g., rheumatoid arthritis), inflammation,  
CC allergies, neurological disorders (e.g., Alzheimer's disease,  
CC Parkinson's disease), cognitive disorders, schizophrenia, asthma,  
CC skin disorders (e.g., psoriasis), sepsis, diabetes, atherosclerosis,  
CC cardiovascular disorders, angioecnic disorders, kidney disorders,  
CC gastrointestinal disorders, pregnancy-related disorders, endocrine  
CC disorders, and infections. The proteins can also be used to aid wound  
CC healing and epithelial cell proliferation, to prevent skin aging due to  
CC sunburn, to maintain organs before transplantation, for supporting cell  
CC culture of primary tissues, to regenerate tissues, to identify their  
CC cognate ligands or binding partners, and in chemotaxis, and can be used



CC as a food additive or preservative to modify storage properties.  
CC Antibodies specific for a protein of the invention can be used in  
CC alleviating symptoms associated with the disorders mentioned above, and  
CC in diagnostic immunoassays e.g., radioimmunoassay or enzyme linked  
CC immunosorbent assay (ELISA). The present sequence represents a human  
CC secreted protein-encoding cDNA of the invention.

XX  
SQ Sequence 347 BP; 73 A; 87 C; 106 G; 76 T; 5 other;

Query Match 90.0%; Score 14.4; DB 22; Length 347;  
Best Local Similarity 93.8%; Pred. No. 3.9e+02;

Matches 15; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 ccttcaccaccccc 16  
||||| |||||||

Db 31 CCTTCTCCACACCCC 16

RESULT 8

AAI83975/c  
ID AAI83975 standard; cDNA: 411 BP.

XX AAI83975;

DT 06-NOV-2001 (first entry)

XX Human polynucleotide SEQ ID NO 4035.

XX Human; cytokine; cell proliferation; cell differentiation; gene therapy;  
KW vaccine; peptide therapy; stem cell growth factor; haematopoiesis;  
KW tissue growth factor; immunomodulatory; cancer; leukaemia;  
KW nervous system disorders; arthritis; inflammation; ss.

XX Homo sapiens.

OS WO200164835-A2.

PN 07-SEP-2001.

PD 26-FEB-2001; 2001WO-US04927.

XX 28-FEB-2000; 2000US-0515126.

PR 18-MAY-2000; 2000US-0577409.

XX (HYSE-) HYSEQ INC.

PI Tang YT, Liu C, Drmanac RT;

XX WPI; 2001-514838/56.

DR P-PSDB; AAO04044.

XX Isolated nucleic acids and polypeptides, useful for preventing  
PT diagnosing and treating e.g. leukaemia, inflammation and immune  
PT disorders -

PS Claim 1; SEQ ID NO 4035; 1399pp + Sequence Listing; English.

CC The invention relates to human polynucleotides (AAI79941-AAI93841) and  
CC the encoded proteins (AAO00010-AAO13910) that exhibit activity elating to  
CC cytokine, cell proliferation or cell differentiation or which may induce  
CC production of other cytokines in other cell populations. The  
CC polynucleotides and polypeptides are useful in gene therapy, vaccines or  
CC peptide therapy. The polypeptides have various cytokine-like activities,  
CC e.g. stem cell growth factor activity, haematopoiesis regulating  
CC activity, tissue growth factor activity, immunomodulatory activity and  
CC activin/inhibin activity and may be useful in the diagnosis and/or  
CC treatment of cancer, leukaemia, nervous system disorders, arthritis and  
CC inflammation.

CC Note: The sequence data for this patent did not form part of the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pct\_sequences.

SQ Sequence 411 BP; 149 A; 73 C; 116 G; 72 T; 1 other;

Query Match 90.0%; Score 14.4; DB 22; Length 411;  
Best Local Similarity 93.8%; Pred. No. 4e+02;

Matches 15; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 ccttcaccaccccc 16  
||||| |||||||

Db 173 CCTTCCACACACCCC 158

RESULT 9

ABAI1306/c  
ID ABAI1306 standard; cDNA: 457 BP.

XX ABAI1306;

DT 23-JAN-2002 (first entry)

XX Human nervous system related polynucleotide SEQ ID NO 313.

XX Human; nootropic; neuroprotective; cytostatic; dermatological; virocidic;  
KW immunosuppressive; antiinflammatory; anti-HIV; antibacterial; vulnerary;  
KW antiparkinsonian; antisclerling; antinaemic; antiarthritis; cancer;  
KW antirheumatic; hepatotropic; cerebroprotective; antiinflammatory;  
KW antiallergic; antidiabetic; antilucer; anticonvulsant; antifungal;  
KW antiparasitic; cardiac; immune disorder; cardiovascular disorder;  
KW neurological disease; infection; neurotropic; gene therapy; vaccine; ss.

XX Homo sapiens.

OS WO200159063-A2.

PN 16-AUG-2001.

PD 17-JAN-2001; 2001WO-US01334.

XX 31-JAN-2000; 2000US-0179065.

PR 04-FEB-2000; 2000US-0180628.

PR 24-FEB-2000; 2000US-0184664.

PR 02-MAR-2000; 2000US-0186350.

PR 16-MAR-2000; 2000US-0189874.

PR 17-MAR-2000; 2000US-0190076.

PR 18-APR-2000; 2000US-0198123.

PR 19-MAY-2000; 2000US-0205515.

PR 07-JUN-2000; 2000US-0209467.

PR 28-JUN-2000; 2000US-0214886.

PR 30-JUN-2000; 2000US-0215135.

PR 07-JUL-2000; 2000US-0216647.

PR 07-JUL-2000; 2000US-0216860.

PR 11-JUL-2000; 2000US-0217487.

PR 11-JUL-2000; 2000US-0217496.

PR 14-JUL-2000; 2000US-0218290.

PR 26-JUL-2000; 2000US-0220963.

PR 26-JUL-2000; 2000US-0220964.

PR 14-AUG-2000; 2000US-0224518.

PR 14-AUG-2000; 2000US-0224519.

PR 14-AUG-2000; 2000US-0225213.

PR 14-AUG-2000; 2000US-0225214.

PR 14-AUG-2000; 2000US-0225266.

PR 14-AUG-2000; 2000US-0225267.

PR 14-AUG-2000; 2000US-0225268.

PR 14-AUG-2000; 2000US-0225270.

PR 14-AUG-2000; 2000US-0225271.

PR 14-AUG-2000; 2000US-0225275.

PR 14-AUG-2000; 2000US-0225758.

PR 14-AUG-2000; 2000US-0225759.

PR 18-AUG-2000; 2000US-0226279.

PR 22-AUG-2000; 2000US-0226681.

PR 22-AUG-2000; 2000US-0226868.

PR 22-AUG-2000; 2000US-0227182.

PR 23-AUG-2000; 2000US-0227009.

PR 30-AUG-2000; 2000US-0228924.  
PR 01-SEP-2000; 2000US-0229287.  
PR 01-SEP-2000; 2000US-0229343.  
PR 01-SEP-2000; 2000US-0229344.  
PR 01-SEP-2000; 2000US-0229345.  
PR 05-SEP-2000; 2000US-0229509.  
PR 05-SEP-2000; 2000US-0229513.  
PR 06-SEP-2000; 2000US-0230437.  
PR 06-SEP-2000; 2000US-0230438.  
PR 08-SEP-2000; 2000US-0231242.  
PR 08-SEP-2000; 2000US-0231243.  
PR 08-SEP-2000; 2000US-0231244.  
PR 08-SEP-2000; 2000US-0231413.  
PR 08-SEP-2000; 2000US-0231414.  
PR 08-SEP-2000; 2000US-0232080.  
PR 08-SEP-2000; 2000US-0232081.  
PR 12-SEP-2000; 2000US-0231968.  
PR 14-SEP-2000; 2000US-0232397.  
PR 14-SEP-2000; 2000US-0232398.  
PR 14-SEP-2000; 2000US-0232399.  
PR 14-SEP-2000; 2000US-0232400.  
PR 14-SEP-2000; 2000US-0232401.  
PR 14-SEP-2000; 2000US-0232063.  
PR 14-SEP-2000; 2000US-0233064.  
PR 14-SEP-2000; 2000US-0233065.  
PR 21-SEP-2000; 2000US-0234223.  
PR 21-SEP-2000; 2000US-0234274.  
PR 25-SEP-2000; 2000US-0234997.  
PR 25-SEP-2000; 2000US-0234998.  
PR 26-SEP-2000; 2000US-0235484.  
PR 27-SEP-2000; 2000US-0235834.  
PR 27-SEP-2000; 2000US-0235836.  
PR 29-SEP-2000; 2000US-0236327.  
PR 29-SEP-2000; 2000US-0236367.  
PR 29-SEP-2000; 2000US-0236368.  
PR 29-SEP-2000; 2000US-0236369.  
PR 29-SEP-2000; 2000US-0236370.  
PR 02-OCT-2000; 2000US-0236802.  
PR 02-OCT-2000; 2000US-0237037.  
PR 02-OCT-2000; 2000US-0237038.  
PR 02-OCT-2000; 2000US-0237039.  
PR 02-OCT-2000; 2000US-0237040.  
PR 13-OCT-2000; 2000US-0239935.  
PR 13-OCT-2000; 2000US-0239937.  
PR 20-OCT-2000; 2000US-0240960.  
PR 20-OCT-2000; 2000US-0241785.  
PR 20-OCT-2000; 2000US-0241786.  
PR 20-OCT-2000; 2000US-0241787.  
PR 20-OCT-2000; 2000US-0241808.  
PR 20-OCT-2000; 2000US-0241809.  
PR 20-OCT-2000; 2000US-0241826.  
PR 01-NOV-2000; 2000US-0242221.  
PR 01-NOV-2000; 2000US-0244617.  
PR 08-NOV-2000; 2000US-0246474.  
PR 08-NOV-2000; 2000US-0246475.  
PR 08-NOV-2000; 2000US-0246476.  
PR 08-NOV-2000; 2000US-0246477.  
PR 08-NOV-2000; 2000US-0246478.  
PR 08-NOV-2000; 2000US-0246523.  
PR 08-NOV-2000; 2000US-0246524.  
PR 08-NOV-2000; 2000US-0246525.  
PR 08-NOV-2000; 2000US-0246526.  
PR 08-NOV-2000; 2000US-0246527.  
PR 08-NOV-2000; 2000US-0246528.  
PR 08-NOV-2000; 2000US-0246532.  
PR 08-NOV-2000; 2000US-0246533.  
PR 08-NOV-2000; 2000US-0246534.  
PR 08-NOV-2000; 2000US-0246535.  
PR 08-NOV-2000; 2000US-0246610.  
PR 08-NOV-2000; 2000US-0246611.  
PR 08-NOV-2000; 2000US-0246613.  
PR 17-NOV-2000; 2000US-0249207.  
PR 17-NOV-2000; 2000US-0249208.  
PR 17-NOV-2000; 2000US-0249209.  
PR 17-NOV-2000; 2000US-0249210.

PR 17-NOV-2000; 2000US-0249211.  
PR 17-NOV-2000; 2000US-0249212.  
PR 17-NOV-2000; 2000US-0249213.  
PR 17-NOV-2000; 2000US-0249214.  
PR 17-NOV-2000; 2000US-0249215.  
PR 17-NOV-2000; 2000US-0249216.  
PR 17-NOV-2000; 2000US-0249217.  
PR 17-NOV-2000; 2000US-0249218.  
PR 17-NOV-2000; 2000US-0249219.  
PR 17-NOV-2000; 2000US-0249224.  
PR 17-NOV-2000; 2000US-0249245.  
PR 17-NOV-2000; 2000US-0249264.  
PR 17-NOV-2000; 2000US-0249265.  
PR 17-NOV-2000; 2000US-0249297.  
PR 17-NOV-2000; 2000US-0249299.  
PR 17-NOV-2000; 2000US-0249300.  
PR 01-DEC-2000; 2000US-0250391.  
PR 01-DEC-2000; 2000US-0251160.  
PR 05-DEC-2000; 2000US-0251030.  
PR 05-DEC-2000; 2000US-0251988.  
PR 05-DEC-2000; 2000US-0256719.  
PR 06-DEC-2000; 2000US-0251479.  
PR 08-DEC-2000; 2000US-0251856.  
PR 08-DEC-2000; 2000US-0251858.  
PR 08-DEC-2000; 2000US-0251869.  
PR 08-DEC-2000; 2000US-0251889.  
PR 08-DEC-2000; 2000US-0251990.  
PR 11-DEC-2000; 2000US-0254097.  
PR 05-JAN-2001; 2001US-0259678.  
XX  
XX  
PA (HUMA-) HUMAN GENOME SCI INC.  
XX  
XX  
PI Rosen CA, Barash SC, Ruben SM;  
XX  
XX MPI: 2001-541565/60.  
DR P-PSDB: ABB14980.  
XX  
XX  
PT Nucleic acids encoding 3224 human nervous system antigen polypeptides,  
PT useful for preventing, diagnosing and/or treating nervous system  
PT cancers and metastases -  
XX  
XX  
PS Claim 1: SEQ ID NO 313; 1701pp + Sequence Listing; English.  
XX  
XX  
CC The invention relates to novel genes (ABA11004-ABA21534) and proteins  
CC (ABA14678-ABA18001) useful for preventing, treating or ameliorating  
CC medical conditions e.g. by protein or gene therapy. The genes are  
CC isolated from a range of human tissues disclosed in the specification.  
CC The nucleic acids, proteins, antibodies and (ant)agonists are useful  
CC in the diagnosis, treatment and prevention of: (a) cancer, e.g. breast  
CC and ovarian cancer and other cancers of the adrenal gland, bone, bone  
CC marrow, breast, gastrointestinal tract, liver, lung, or urogenital;  
CC (b) immune disorders e.g. Addison's disease, allergies, autoimmune  
CC haemolytic anaemia, autoimmune thyroiditis, diabetes mellitus, Crohn's  
CC disease, multiple sclerosis, rheumatoid arthritis and ulcerative  
CC colitis; (c) cardiovascular disorders such as myocardial ischaemia;  
CC (d) wound healing; (e) neurological diseases e.g. cerebral anoxia and  
CC epilepsy; and (f) infectious diseases such as viral, bacterial, fungal  
CC and parasitic infections.  
CC Note: The sequence data for this patent did not form part of the  
CC printed specification, but was obtained in electronic format directly  
CC from Wipo at ftp.wipo.int/pub/published\_pcl\_sequences.  
XX  
XX  
SQ Sequence 457 BP; 141 A; 81 C; 86 G; 144 T; 5 other:  
  
Query Match 90.0%; Score 14.4; DB 22; Length 457;  
Best Local Similarity 93.8%; Pred. No. 4e+02;  
Matches 15; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
  
QY 1 ccttaccacacccc 16  
DB 457 CCTTACCAACACC 442

RESULT 10  
AAS64618/c  
ID AAS64618 standard; cDNA; 606 BP.  
XX  
XX AAS64618;  
XX  
XX 13-FEB-2002 (first entry)  
XX  
XX DNA encoding novel human diagnostic protein #422.  
DE  
XX Human: chromosome mapping; gene mapping; gene therapy; forensic;  
KW food supplement; medical imaging; diagnostic; genetic disorder; ss.  
XX  
XX Homo sapiens.  
OS  
XX WO200175067-A2.  
PN  
XX 11-OCT-2001.  
PD  
XX 30-MAR-2001; 2001WO-US08631.  
PF  
XX 31-MAR-2000; 2000US-0540217.  
PR 23-AUG-2000; 2000US-0649167.  
XX  
XX (HYSE-) HYSEQ INC.  
PA  
XX Drmanac RT, Liu C, Tang YT;  
PI  
XX WPI; 2001-639362/73.  
DR P-PSDB; ABG00431.  
DR  
XX  
XX New isolated polynucleotide and encoded polypeptides, useful in  
PT diagnostics, forensics, gene mapping, identification of mutations  
PT responsible for genetic disorders or other traits and to assess  
PT biodiversity  
PT  
XX  
XX Claim 1; SEQ ID No 422; 103pp; English.  
PS  
XX  
XX The invention relates to isolated polynucleotide (I) and  
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,  
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome  
CC and gene mapping, and in recombinant production of (II). The  
CC polynucleotides are also used in diagnostics as expressed sequence tags  
CC for identifying expressed genes. (I) is useful in gene therapy techniques  
CC to restore normal activity of (II) or to treat disease states involving  
CC (II). (II) is useful for generating antibodies against it, detecting or  
CC quantitating a polypeptide in tissue, as molecular weight markers and as  
CC a food supplement. (II) and its binding partners are useful in medical  
CC imaging of sites expressing (II). (I) and (II) are useful for treating  
CC disorders involving aberrant protein expression or biological activity.  
CC The polypeptide and polynucleotide sequences have applications in  
CC diagnostics, forensics, gene mapping, identification of mutations  
CC responsible for genetic disorders or other traits to assess biodiversity  
CC and to produce other types of data and products dependent on DNA and  
CC amino acid sequences. AAS64197-AAS94564 represent novel human  
CC diagnostic coding sequences of the invention.  
CC Note: The sequence data for this patent did not appear in the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pct\_sequences.  
XX  
SQ Sequence 606 BP; 128 A; 188 C; 133 G; 157 T; 0 other;

Query Match 90.0%; Score 14.4; DB 23; Length 606;  
Best Local Similarity 93.8%; Pred. No. 4e+02; 1; Indels 0; Gaps 0;  
Matches 15; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 ccttcaccacccccc 16  
|||||  
Db 532 CCTTCAACCAATCCCC 517

RESULT 11

AAK70540/c  
ID AAK70540 standard; DNA; 736 BP.  
XX  
XX AAK70540;  
AC  
XX 06-NOV-2001 (first entry)  
DT  
XX  
XX Human immune/haematopoietic antigen genomic sequence SEQ ID NO:25352.  
DE  
XX  
XX Human: immune; haematopoietic; immune/haematopoietic antigen; cancer;  
KW cytostatic; gene therapy; vaccine; metastasis; ds.  
XX  
XX Homo sapiens.  
OS  
XX WO200157182-A2.  
PN  
XX 09-AUG-2001.  
PD  
XX  
XX 17-JAN-2001; 2001WO-US01354.  
PF  
XX 31-JAN-2000; 2000US-0179065.  
PR 04-FEB-2000; 2000US-0180628.  
PR 24-FEB-2000; 2000US-0184664.  
PR 02-MAR-2000; 2000US-0186350.  
PR 16-MAR-2000; 2000US-0189874.  
PR 17-MAR-2000; 2000US-0190076.  
PR 18-APR-2000; 2000US-0198123.  
PR 19-MAY-2000; 2000US-0205515.  
PR 07-JUN-2000; 2000US-0209467.  
PR 28-JUN-2000; 2000US-0214886.  
PR 30-JUN-2000; 2000US-0215135.  
PR 07-JUL-2000; 2000US-0216647.  
PR 07-JUL-2000; 2000US-0216880.  
PR 11-JUL-2000; 2000US-0217487.  
PR 11-JUL-2000; 2000US-0217496.  
PR 14-JUL-2000; 2000US-0218290.  
PR 26-JUL-2000; 2000US-0220963.  
PR 26-JUL-2000; 2000US-0220964.  
PR 14-AUG-2000; 2000US-0224518.  
PR 14-AUG-2000; 2000US-0224519.  
PR 14-AUG-2000; 2000US-0225213.  
PR 14-AUG-2000; 2000US-0225214.  
PR 14-AUG-2000; 2000US-0225266.  
PR 14-AUG-2000; 2000US-0225267.  
PR 14-AUG-2000; 2000US-0225268.  
PR 14-AUG-2000; 2000US-0225270.  
PR 14-AUG-2000; 2000US-0225447.  
PR 14-AUG-2000; 2000US-0225757.  
PR 14-AUG-2000; 2000US-0225758.  
PR 14-AUG-2000; 2000US-0225759.  
PR 18-AUG-2000; 2000US-0226279.  
PR 22-AUG-2000; 2000US-0226681.  
PR 22-AUG-2000; 2000US-0226868.  
PR 22-AUG-2000; 2000US-0227182.  
PR 23-AUG-2000; 2000US-0228924.  
PR 30-AUG-2000; 2000US-0228924.  
PR 01-SEP-2000; 2000US-0229287.  
PR 01-SEP-2000; 2000US-0229343.  
PR 01-SEP-2000; 2000US-0229344.  
PR 01-SEP-2000; 2000US-0229345.  
PR 01-SEP-2000; 2000US-0229509.  
PR 05-SEP-2000; 2000US-0229513.  
PR 06-SEP-2000; 2000US-0230437.  
PR 06-SEP-2000; 2000US-0230438.  
PR 08-SEP-2000; 2000US-0231242.  
PR 08-SEP-2000; 2000US-0231243.  
PR 08-SEP-2000; 2000US-0231244.  
PR 08-SEP-2000; 2000US-0231413.  
PR 08-SEP-2000; 2000US-0231414.  
PR 08-SEP-2000; 2000US-0232080.  
PR 08-SEP-2000; 2000US-0232081.  
PR 12-SEP-2000; 2000US-0231968.  
PR 14-SEP-2000; 2000US-0232397.

PR 14-SEP-2000; 2000US-0232398.  
 PR 14-SEP-2000; 2000US-0232399.  
 PR 14-SEP-2000; 2000US-0232400.  
 PR 14-SEP-2000; 2000US-0232401.  
 PR 14-SEP-2000; 2000US-0233065.  
 PR 14-SEP-2000; 2000US-0233066.  
 PR 14-SEP-2000; 2000US-0233067.  
 PR 21-SEP-2000; 2000US-0234223.  
 PR 21-SEP-2000; 2000US-0234224.  
 PR 25-SEP-2000; 2000US-0234997.  
 PR 25-SEP-2000; 2000US-0234998.  
 PR 26-SEP-2000; 2000US-0235484.  
 PR 27-SEP-2000; 2000US-0235834.  
 PR 29-SEP-2000; 2000US-0236327.  
 PR 29-SEP-2000; 2000US-0236367.  
 PR 29-SEP-2000; 2000US-0236368.  
 PR 29-SEP-2000; 2000US-0236369.  
 PR 29-SEP-2000; 2000US-0236370.  
 PR 02-OCT-2000; 2000US-0236802.  
 PR 02-OCT-2000; 2000US-0237037.  
 PR 02-OCT-2000; 2000US-0237038.  
 PR 02-OCT-2000; 2000US-0237039.  
 PR 02-OCT-2000; 2000US-0237040.  
 PR 13-OCT-2000; 2000US-0239935.  
 PR 13-OCT-2000; 2000US-0239937.  
 PR 20-OCT-2000; 2000US-0240960.  
 PR 20-OCT-2000; 2000US-0241221.  
 PR 20-OCT-2000; 2000US-0241785.  
 PR 20-OCT-2000; 2000US-0241786.  
 PR 20-OCT-2000; 2000US-0241787.  
 PR 20-OCT-2000; 2000US-0241808.  
 PR 20-OCT-2000; 2000US-0241809.  
 PR 20-OCT-2000; 2000US-0241826.  
 PR 01-NOV-2000; 2000US-0244617.  
 PR 08-NOV-2000; 2000US-0246474.  
 PR 08-NOV-2000; 2000US-0246475.  
 PR 08-NOV-2000; 2000US-0246476.  
 PR 08-NOV-2000; 2000US-0246477.  
 PR 08-NOV-2000; 2000US-0246478.  
 PR 08-NOV-2000; 2000US-0246523.  
 PR 08-NOV-2000; 2000US-0246524.  
 PR 08-NOV-2000; 2000US-0246525.  
 PR 08-NOV-2000; 2000US-0246526.  
 PR 08-NOV-2000; 2000US-0246527.  
 PR 08-NOV-2000; 2000US-0246528.  
 PR 08-NOV-2000; 2000US-0246532.  
 PR 08-NOV-2000; 2000US-0246609.  
 PR 08-NOV-2000; 2000US-0246610.  
 PR 08-NOV-2000; 2000US-0246611.  
 PR 08-NOV-2000; 2000US-0246613.  
 PR 17-NOV-2000; 2000US-0249207.  
 PR 17-NOV-2000; 2000US-0249208.  
 PR 17-NOV-2000; 2000US-0249209.  
 PR 17-NOV-2000; 2000US-0249210.  
 PR 17-NOV-2000; 2000US-0249211.  
 PR 17-NOV-2000; 2000US-0249212.  
 PR 17-NOV-2000; 2000US-0249213.  
 PR 17-NOV-2000; 2000US-0249214.  
 PR 17-NOV-2000; 2000US-0249215.  
 PR 17-NOV-2000; 2000US-0249216.  
 PR 17-NOV-2000; 2000US-0249217.  
 PR 17-NOV-2000; 2000US-0249218.  
 PR 17-NOV-2000; 2000US-0249244.  
 PR 17-NOV-2000; 2000US-0249245.  
 PR 17-NOV-2000; 2000US-0249264.  
 PR 17-NOV-2000; 2000US-0249265.  
 PR 17-NOV-2000; 2000US-0249297.  
 PR 17-NOV-2000; 2000US-0249309.  
 PR 17-NOV-2000; 2000US-0249300.  
 PR 01-DEC-2000; 2000US-0250160.  
 PR 01-DEC-2000; 2000US-0250391.  
 PR 05-DEC-2000; 2000US-0251030.

PR 05-DEC-2000; 2000US-0251988.  
 PR 05-DEC-2000; 2000US-0256719.  
 PR 05-DEC-2000; 2000US-0251479.  
 PR 08-DEC-2000; 2000US-0251856.  
 PR 08-DEC-2000; 2000US-0251858.  
 PR 08-DEC-2000; 2000US-0251869.  
 PR 08-DEC-2000; 2000US-0251989.  
 PR 08-DEC-2000; 2000US-0251990.  
 PR 11-DEC-2000; 2000US-0254097.  
 PR 05-JAN-2001; 2001US-0259678.  
 XX  
 PA (HUMA-) HUMAN GENOME SCI INC.  
 XX  
 PI Rosen CA, Barash SC, Ruben SM;  
 XX  
 DR WPI: 2001-483426/52.  
 XX  
 XX  
 PT Nucleic acids encoding human immune/hematopoietic antigen polypeptides,  
 PT useful for preventing, diagnosing and/or treating cancers and  
 PT metastasis -  
 XX  
 PS Disclosure: SEQ ID NO 25352; 3071bp + Sequence listing; English.  
 XX  
 CC AAK54951 to AAK64702 encode the human immune/hematopoietic antigen (I)  
 CC amino acid sequences given in AAK82170 to AAK91921. (I) have cytosolic  
 CC activity, and can be used in gene therapy and vaccine production. (I)  
 CC proteins and polynucleotides may be used in the prevention, diagnosis and  
 CC treatment of diseases associated with inappropriate (I) expression. For  
 CC example, they may be used to treat disorders associated with decreased  
 CC expression by rectifying mutations or deletions in a patient's genome  
 CC that affect the activity of (I) by expressing inactive proteins or to  
 CC supplement the patients own production of (I). Additionally, (I)  
 CC polynucleotides may be used to produce the secreted (I), by inserting the  
 CC the nucleic acids into a host cell and culturing the cell to express the  
 CC protein. (I) proteins and polynucleotides may be used to prevent,  
 CC diagnose and treat immune/hematopoietic-related diseases, especially  
 CC cancers and cancer metastases of hematopoietic-derived cells. AAK64703  
 CC to AAK87694 represent human immune/hematopoietic antigen genomic  
 CC sequences from the present invention. AAK54942 to AAK54950 and AAK82169  
 CC represent sequences used in the exemplification of the present invention.  
 XX  
 SQ Sequence 736 BP; 125 A; 218 C; 192 G; 201 T; 0 other;

Query Match 90.0%; Score 14.4; DB 22; Length 736;  
 Best Local Similarity 93.8%; Pred. No. 4.1e+02;  
 Matches 15; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 ccttcaccaccccc 16  
 ||||| ||||| |||||  
 DB 571 CCTTCCCAACCCCC 556

RESULT 12  
 AAK70541/c  
 ID AAK70541 standard; DNA; 736 BP.  
 XX  
 AC AAK70541;  
 XX  
 DT 06-NOV-2001 (first entry)  
 XX  
 DE Human immune/hematopoietic antigen genomic sequence SEQ ID NO:25353.  
 XX  
 XX Human; immune; haematopoietic; immune/hematopoietic antigen; cancer;  
 KM cytosolic; gene therapy; vaccine; metastasis; ds.  
 XX  
 OS Homo sapiens.  
 OS  
 PN WO200157182-A2.  
 XX  
 PD 09-AUG-2001.  
 PD  
 XX 17-JAN-2001; 2001WO-US01354.  
 PF

XX 31-JAN-2000; 2000US-0179065.  
PR 04-FEB-2000; 2000US-0180628.  
PR 24-FEB-2000; 2000US-0184664.  
PR 02-MAR-2000; 2000US-0186350.  
PR 16-MAR-2000; 2000US-0189874.  
PR 17-MAR-2000; 2000US-0190076.  
PR 18-APR-2000; 2000US-0198123.  
PR 19-MAY-2000; 2000US-0205515.  
PR 07-JUN-2000; 2000US-0209467.  
PR 28-JUN-2000; 2000US-0214886.  
PR 30-JUN-2000; 2000US-0215135.  
PR 07-JUL-2000; 2000US-0216647.  
PR 11-JUL-2000; 2000US-0216880.  
PR 11-JUL-2000; 2000US-0217487.  
PR 11-JUL-2000; 2000US-0217496.  
PR 14-JUL-2000; 2000US-0218290.  
PR 26-JUL-2000; 2000US-0220963.  
PR 26-JUL-2000; 2000US-0220963.  
PR 14-AUG-2000; 2000US-0224518.  
PR 14-AUG-2000; 2000US-0224519.  
PR 14-AUG-2000; 2000US-0225213.  
PR 14-AUG-2000; 2000US-0225214.  
PR 14-AUG-2000; 2000US-0225266.  
PR 14-AUG-2000; 2000US-0225267.  
PR 14-AUG-2000; 2000US-0225268.  
PR 14-AUG-2000; 2000US-0225270.  
PR 14-AUG-2000; 2000US-0225447.  
PR 14-AUG-2000; 2000US-0225757.  
PR 14-AUG-2000; 2000US-0225758.  
PR 14-AUG-2000; 2000US-0225759.  
PR 18-AUG-2000; 2000US-0226279.  
PR 22-AUG-2000; 2000US-0226681.  
PR 22-AUG-2000; 2000US-0226688.  
PR 22-AUG-2000; 2000US-0227182.  
PR 23-AUG-2000; 2000US-0227009.  
PR 30-AUG-2000; 2000US-0228924.  
PR 01-SEP-2000; 2000US-0229287.  
PR 01-SEP-2000; 2000US-0229343.  
PR 01-SEP-2000; 2000US-0229344.  
PR 01-SEP-2000; 2000US-0229345.  
PR 05-SEP-2000; 2000US-0229309.  
PR 05-SEP-2000; 2000US-0229513.  
PR 06-SEP-2000; 2000US-0230437.  
PR 06-SEP-2000; 2000US-0230438.  
PR 08-SEP-2000; 2000US-0231242.  
PR 08-SEP-2000; 2000US-0231243.  
PR 08-SEP-2000; 2000US-0231244.  
PR 08-SEP-2000; 2000US-0231413.  
PR 08-SEP-2000; 2000US-0231414.  
PR 08-SEP-2000; 2000US-0232080.  
PR 08-SEP-2000; 2000US-0232081.  
PR 12-SEP-2000; 2000US-0231968.  
PR 14-SEP-2000; 2000US-0232397.  
PR 14-SEP-2000; 2000US-0232398.  
PR 14-SEP-2000; 2000US-0232399.  
PR 14-SEP-2000; 2000US-0232400.  
PR 14-SEP-2000; 2000US-0232401.  
PR 14-SEP-2000; 2000US-0233063.  
PR 14-SEP-2000; 2000US-0233064.  
PR 14-SEP-2000; 2000US-0233065.  
PR 21-SEP-2000; 2000US-0234223.  
PR 21-SEP-2000; 2000US-0234274.  
PR 25-SEP-2000; 2000US-0234997.  
PR 25-SEP-2000; 2000US-0234998.  
PR 26-SEP-2000; 2000US-0235484.  
PR 27-SEP-2000; 2000US-0235834.  
PR 27-SEP-2000; 2000US-0235836.  
PR 29-SEP-2000; 2000US-0236327.  
PR 29-SEP-2000; 2000US-0236367.  
PR 29-SEP-2000; 2000US-0236368.  
PR 29-SEP-2000; 2000US-0236369.  
PR 29-SEP-2000; 2000US-0236370.

PR 02-OCT-2000; 2000US-0236802.  
PR 02-OCT-2000; 2000US-0237037.  
PR 02-OCT-2000; 2000US-0237038.  
PR 02-OCT-2000; 2000US-0237039.  
PR 02-OCT-2000; 2000US-0237040.  
PR 13-OCT-2000; 2000US-0239935.  
PR 13-OCT-2000; 2000US-0239937.  
PR 20-OCT-2000; 2000US-0240960.  
PR 20-OCT-2000; 2000US-0241221.  
PR 20-OCT-2000; 2000US-0241785.  
PR 20-OCT-2000; 2000US-0241786.  
PR 20-OCT-2000; 2000US-0241787.  
PR 20-OCT-2000; 2000US-0241808.  
PR 20-OCT-2000; 2000US-0241809.  
PR 20-OCT-2000; 2000US-0241826.  
PR 01-NOV-2000; 2000US-0244617.  
PR 01-NOV-2000; 2000US-0244617.  
PR 08-NOV-2000; 2000US-0246474.  
PR 08-NOV-2000; 2000US-0246475.  
PR 08-NOV-2000; 2000US-0246476.  
PR 08-NOV-2000; 2000US-0246477.  
PR 08-NOV-2000; 2000US-0246478.  
PR 08-NOV-2000; 2000US-0246523.  
PR 08-NOV-2000; 2000US-0246524.  
PR 08-NOV-2000; 2000US-0246525.  
PR 08-NOV-2000; 2000US-0246526.  
PR 08-NOV-2000; 2000US-0246527.  
PR 08-NOV-2000; 2000US-0246528.  
PR 08-NOV-2000; 2000US-0246532.  
PR 08-NOV-2000; 2000US-0246609.  
PR 08-NOV-2000; 2000US-0246610.  
PR 08-NOV-2000; 2000US-0246611.  
PR 08-NOV-2000; 2000US-0246613.  
PR 17-NOV-2000; 2000US-0249207.  
PR 17-NOV-2000; 2000US-0249208.  
PR 17-NOV-2000; 2000US-0249209.  
PR 17-NOV-2000; 2000US-0249210.  
PR 17-NOV-2000; 2000US-0249211.  
PR 17-NOV-2000; 2000US-0249212.  
PR 17-NOV-2000; 2000US-0249213.  
PR 17-NOV-2000; 2000US-0249214.  
PR 17-NOV-2000; 2000US-0249215.  
PR 17-NOV-2000; 2000US-0249216.  
PR 17-NOV-2000; 2000US-0249217.  
PR 17-NOV-2000; 2000US-0249218.  
PR 17-NOV-2000; 2000US-0249244.  
PR 17-NOV-2000; 2000US-0249245.  
PR 17-NOV-2000; 2000US-0249264.  
PR 17-NOV-2000; 2000US-0249265.  
PR 17-NOV-2000; 2000US-0249297.  
PR 17-NOV-2000; 2000US-0249299.  
PR 17-NOV-2000; 2000US-0249300.  
PR 01-DEC-2000; 2000US-0250160.  
PR 01-DEC-2000; 2000US-0250391.  
PR 05-DEC-2000; 2000US-0251030.  
PR 05-DEC-2000; 2000US-0251988.  
PR 06-DEC-2000; 2000US-0256719.  
PR 06-DEC-2000; 2000US-0251479.  
PR 08-DEC-2000; 2000US-0251856.  
PR 08-DEC-2000; 2000US-0251868.  
PR 08-DEC-2000; 2000US-0251869.  
PR 08-DEC-2000; 2000US-0251869.  
PR 08-DEC-2000; 2000US-0251889.  
PR 11-DEC-2000; 2000US-0251990.  
PR 05-JAN-2001; 2001US-0259678.

(HUMA-) HUMAN GENOME SCI INC.

Rosen CA, Barash SC, Ruben SM;

WPI, 2001-483426/52.

Nucleic acids encoding human immune/hematopoietic antigen polypeptides,  
useful for preventing, diagnosing and/or treating cancers and

PT metastasis -  
XX Disclosure; SEQ ID NO 25353; 3071pp + Sequence Listing; English.  
XX  
CC AAK54951 to AAK64702 encode the human immune/haematopoietic antigen (I)  
CC amino acid sequences given in AAM82170 to AAM91921. (I) have cytostatic  
CC activity, and can be used in gene therapy and vaccine production. (I)  
CC proteins and polynucleotides may be used in the prevention, diagnosis and  
CC treatment of diseases associated with inappropriate (I) expression. For  
CC example, they may be used to treat disorders associated with decreased  
CC expression by rectifying mutations or deletions in a patient's genome  
CC that affect the activity of (I) by expressing inactive proteins or to  
CC supplement the patient's own production of (I). Additionally, (I)  
CC polynucleotides may be used to produce the secreted (I), by inserting  
CC the nucleic acids into a host cell and culturing the cell to express the  
CC protein. (I) proteins and polynucleotides may be used to prevent,  
CC diagnose and treat immune/haematopoietic-related diseases, especially  
CC cancers and cancer metastases of haematopoietic-derived cells. AAK64703  
CC to AAK87694 represent human immune/haematopoietic antigen genomic  
CC sequences from the present invention. AAK54942 to AAK54950 and AAM82169  
CC represent sequences used in the exemplification of the present invention.  
XX  
SQ Sequence 736 BP; 126 A; 218 C; 191 G; 201 T; 0 other;

Query Match 90.0%; Score 14.4; DB 22; Length 736;  
Best Local Similarity 93.8%; Pred. No. 4.1e+02;  
Matches 15; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 cctttaccaccccc 16  
||||| |||||||  
Db 571 CcTTCCcCAACCCc 556

RESULT 13  
AAK70542/C  
ID AAK70542 standard; DNA: 736 BP.

XX AAK70542;

DT 06-NOV-2001 (first entry)

DE Human immune/haematopoietic antigen genomic sequence SEQ ID NO:25354.

XX Human; immune; haematopoietic; immune/haematopoietic antigen; cancer;

KM cytostatic; gene therapy; vaccine; metastasis; ds.

XX Homo sapiens.

PN WO200157182-A2.

XX 09-AUG-2001.

PD 17-JAN-2001; 2001WO-US01354.

XX 31-JAN-2000; 2000US-0179065.

PR 04-FEB-2000; 2000US-0180628.

PR 24-FEB-2000; 2000US-0184664.

PR 02-MAR-2000; 2000US-0186350.

PR 16-MAR-2000; 2000US-0189874.

PR 17-MAR-2000; 2000US-0190076.

PR 18-APR-2000; 2000US-0198123.

PR 19-MAY-2000; 2000US-0205515.

PR 07-JUN-2000; 2000US-0209467.

PR 28-JUN-2000; 2000US-0214886.

PR 30-JUN-2000; 2000US-0215135.

PR 07-JUL-2000; 2000US-0216647.

PR 11-JUL-2000; 2000US-0217487.

PR 14-JUL-2000; 2000US-0217496.

PR 26-JUL-2000; 2000US-0220963.

PR 26-JUL-2000; 2000US-0220964.

PR 14-AUG-2000; 2000US-0224518.  
PR 14-AUG-2000; 2000US-0224519.  
PR 14-AUG-2000; 2000US-0225213.  
PR 14-AUG-2000; 2000US-0225214.  
PR 14-AUG-2000; 2000US-0225266.  
PR 14-AUG-2000; 2000US-0225267.  
PR 14-AUG-2000; 2000US-0225268.  
PR 14-AUG-2000; 2000US-0225270.  
PR 14-AUG-2000; 2000US-0225447.  
PR 14-AUG-2000; 2000US-0225757.  
PR 14-AUG-2000; 2000US-0225758.  
PR 14-AUG-2000; 2000US-0225759.  
PR 18-AUG-2000; 2000US-0226279.  
PR 22-AUG-2000; 2000US-0226681.  
PR 22-AUG-2000; 2000US-0226688.  
PR 23-AUG-2000; 2000US-0227182.  
PR 30-AUG-2000; 2000US-0227009.  
PR 01-SEP-2000; 2000US-0228924.  
PR 01-SEP-2000; 2000US-0229287.  
PR 01-SEP-2000; 2000US-0229343.  
PR 01-SEP-2000; 2000US-0229344.  
PR 01-SEP-2000; 2000US-0229345.  
PR 05-SEP-2000; 2000US-0229509.  
PR 06-SEP-2000; 2000US-0229513.  
PR 06-SEP-2000; 2000US-0230437.  
PR 08-SEP-2000; 2000US-0230438.  
PR 08-SEP-2000; 2000US-0231242.  
PR 08-SEP-2000; 2000US-0231243.  
PR 08-SEP-2000; 2000US-0231244.  
PR 08-SEP-2000; 2000US-0231413.  
PR 08-SEP-2000; 2000US-0231414.  
PR 08-SEP-2000; 2000US-0232080.  
PR 08-SEP-2000; 2000US-0232081.  
PR 12-SEP-2000; 2000US-0231968.  
PR 14-SEP-2000; 2000US-0232397.  
PR 14-SEP-2000; 2000US-0232398.  
PR 14-SEP-2000; 2000US-0232399.  
PR 14-SEP-2000; 2000US-0232400.  
PR 14-SEP-2000; 2000US-0232401.  
PR 14-SEP-2000; 2000US-0233063.  
PR 14-SEP-2000; 2000US-0233064.  
PR 14-SEP-2000; 2000US-0233065.  
PR 21-SEP-2000; 2000US-0234223.  
PR 21-SEP-2000; 2000US-0234274.  
PR 25-SEP-2000; 2000US-0234997.  
PR 25-SEP-2000; 2000US-0234998.  
PR 26-SEP-2000; 2000US-0235484.  
PR 27-SEP-2000; 2000US-0235834.  
PR 27-SEP-2000; 2000US-0235836.  
PR 29-SEP-2000; 2000US-0236327.  
PR 29-SEP-2000; 2000US-0236327.  
PR 29-SEP-2000; 2000US-0236367.  
PR 29-SEP-2000; 2000US-0236368.  
PR 29-SEP-2000; 2000US-0236369.  
PR 29-SEP-2000; 2000US-0236370.  
PR 02-OCT-2000; 2000US-0236802.  
PR 02-OCT-2000; 2000US-0237037.  
PR 02-OCT-2000; 2000US-0237038.  
PR 02-OCT-2000; 2000US-0237039.  
PR 02-OCT-2000; 2000US-0237040.  
PR 13-OCT-2000; 2000US-0239335.  
PR 13-OCT-2000; 2000US-0239337.  
PR 20-OCT-2000; 2000US-0240960.  
PR 20-OCT-2000; 2000US-0241221.  
PR 20-OCT-2000; 2000US-0241785.  
PR 20-OCT-2000; 2000US-0241786.  
PR 20-OCT-2000; 2000US-0241787.  
PR 20-OCT-2000; 2000US-0241808.  
PR 20-OCT-2000; 2000US-0241809.  
PR 20-OCT-2000; 2000US-0244826.  
PR 01-NOV-2000; 2000US-0244617.  
PR 08-NOV-2000; 2000US-0246474.  
PR 08-NOV-2000; 2000US-0246475.  
PR 08-NOV-2000; 2000US-0246476.

PR 08-NOV-2000; 2000US-0246477.  
 PR 08-NOV-2000; 2000US-0246478.  
 PR 08-NOV-2000; 2000US-0246523.  
 PR 08-NOV-2000; 2000US-0246524.  
 PR 08-NOV-2000; 2000US-0246525.  
 PR 08-NOV-2000; 2000US-0246526.  
 PR 08-NOV-2000; 2000US-0246527.  
 PR 08-NOV-2000; 2000US-0246528.  
 PR 08-NOV-2000; 2000US-0246532.  
 PR 08-NOV-2000; 2000US-0246609.  
 PR 08-NOV-2000; 2000US-0246610.  
 PR 08-NOV-2000; 2000US-0246611.  
 PR 08-NOV-2000; 2000US-0246613.  
 PR 17-NOV-2000; 2000US-0249207.  
 PR 17-NOV-2000; 2000US-0249208.  
 PR 17-NOV-2000; 2000US-0249209.  
 PR 17-NOV-2000; 2000US-0249210.  
 PR 17-NOV-2000; 2000US-0249211.  
 PR 17-NOV-2000; 2000US-0249212.  
 PR 17-NOV-2000; 2000US-0249213.  
 PR 17-NOV-2000; 2000US-0249214.  
 PR 17-NOV-2000; 2000US-0249215.  
 PR 17-NOV-2000; 2000US-0249216.  
 PR 17-NOV-2000; 2000US-0249217.  
 PR 17-NOV-2000; 2000US-0249218.  
 PR 17-NOV-2000; 2000US-0249244.  
 PR 17-NOV-2000; 2000US-0249245.  
 PR 17-NOV-2000; 2000US-0249264.  
 PR 17-NOV-2000; 2000US-0249265.  
 PR 17-NOV-2000; 2000US-0249297.  
 PR 17-NOV-2000; 2000US-0249299.  
 PR 17-NOV-2000; 2000US-0249300.  
 PR 01-DEC-2000; 2000US-0250160.  
 PR 01-DEC-2000; 2000US-0250391.  
 PR 05-DEC-2000; 2000US-0251030.  
 PR 05-DEC-2000; 2000US-0251988.  
 PR 05-DEC-2000; 2000US-0256719.  
 PR 06-DEC-2000; 2000US-0251479.  
 PR 08-DEC-2000; 2000US-0251856.  
 PR 08-DEC-2000; 2000US-0251868.  
 PR 08-DEC-2000; 2000US-0251869.  
 PR 08-DEC-2000; 2000US-0251989.  
 PR 08-DEC-2000; 2000US-0251990.  
 PR 11-DEC-2000; 2000US-0254097.  
 PR 05-JAN-2001; 2001US-0259678.  
 PA (HUMA-) HUMAN GENOME SCI INC.  
 PI Rosen CA, Barash SC, Ruben SM;  
 XX WPI; 2001-483426/52.  
 DR Nucleic acids encoding human immune/hematopoietic antigen polypeptides,  
 PT useful for preventing, diagnosing and/or treating cancers and  
 PR metastasis -  
 XX  
 PS Disclosure: SEQ ID NO 25354; 3071bp + Sequence Listing; English.  
 XX  
 XX AAK54951 to AAK64702 encode the human immune/haematopoietic antigen (I)  
 CC amino acid sequences given in AAM82170 to AAM91921. (I) have cytostatic  
 CC activity, and can be used in gene therapy and vaccine production. (I)  
 CC proteins and polynucleotides may be used in the prevention, diagnosis and  
 CC treatment of diseases associated with inappropriate (I) expression. For  
 CC example, they may be used to treat disorders associated with decreased  
 CC expression by rectifying mutations or deletions in a patient's genome  
 CC that affect the activity of (I) by expressing inactive proteins or to  
 CC supplement the patients own production of (I). Additionally, (I)  
 CC polynucleotides may be used to produce the secreted (I), by inserting  
 CC the nucleic acids into a host cell and culturing the cell to express the  
 CC protein. (I) proteins and polynucleotides may be used to prevent,  
 CC diagnose and treat immune/haematopoietic-related diseases, especially  
 CC cancers and cancer metastases of haematopoietic-derived cells. AAK64703  
 CC to AAK87694 represent human immune/haematopoietic antigen genomic

CC sequences from the present invention. AAK54942 to AAK54950 and AAM82169  
 CC represent sequences used in the exemplification of the present invention.  
 XX  
 SQ Sequence 736 BP; 126 A; 212 C; 193 G; 205 T; 0 other;

Query Match 90.0%; Score 14.4; DB 22; Length 736;  
 Best Local Similarity 93.8%; Pred. No. 4.1e+02;  
 Matches 15; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 ccttcaccaaccccc 16  
 Db 571 CcTTCcCAACCCCC 556

RESULT 14  
 AAL15377  
 ID AAL15377 standard; cDNA; 763 BP.  
 XX  
 AC AAL15377;  
 XX  
 DT 07-DEC-2001 (first entry)  
 XX  
 DE Human breast cancer expressed polynucleotide 7834.  
 XX  
 KW Human; breast cancer; cell marker; cytostatic; ss.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO200151628-A2.  
 XX  
 PD 19-JUL-2001.  
 XX  
 PE 10-JAN-2001; 2001MO-US00798.  
 XX  
 PR 14-MAR-2000; 2000US-0176077.  
 PR 14-MAR-2000; 2000US-0189167.  
 PR 24-MAR-2000; 2000US-0192099.  
 PR 29-MAR-2000; 2000US-0193480.  
 PR 15-MAY-2000; 2000US-0205230.  
 PR 09-JUN-2000; 2000US-0211315.  
 PR 25-JUL-2000; 2000US-0220534.  
 XX  
 PA (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.  
 PI Lillie J, Xu Y, Wang Y, Steinmann K;  
 PT WPI; 2001-451856/48.  
 DR  
 XX  
 XX New peptide useful as a marker for the diagnosis of breast cancer -  
 PT  
 PS Claim 1; Page 1411; 3695pp; English.  
 XX  
 XX The invention relates to human breast cancer expressed polynucleotides  
 CC afflicted with breast cancer and methods of assessing whether a patient is  
 CC expression of certain markers and the cancerous state of breast cells.  
 CC The polynucleotides and encoded polypeptides are potential markers for  
 CC detecting, diagnosing, monitoring, characterising treating and  
 CC potentially preventing breast cancer. The polynucleotides and encoded  
 CC polypeptides are also useful for isolating compounds with cytostatic  
 CC activity.  
 XX  
 SQ Sequence 763 BP; 178 A; 183 C; 219 G; 153 T; 30 other;

Query Match 90.0%; Score 14.4; DB 22; Length 763;  
 Best Local Similarity 93.8%; Pred. No. 4.1e+02;  
 Matches 15; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
 QY 1 ccttcaccaaccccc 16  
 Db 1 ccttcaccaaccccc 16

Db 555 CCTTTCATCAACCCC 540

Search completed: July 30, 2002, 00:01:18  
Job time: 4929 sec

RESULT 15  
 ID AAH07529/C  
 ID AAH07529 standard; cDNA; 807 BP.  
 XX  
 AC AAH07529;  
 XX  
 DT 26-JUN-2001 (first entry)  
 XX  
 DE Human cDNA clone (5'-primer) SEQ ID NO:4364.  
 XX  
 KW Human; primer; detection; diagnosis; antisense therapy; gene therapy; ss  
 XX  
 OS Homo sapiens.  
 XX  
 PN EPI074617-A2.  
 XX  
 PD 07-FEB-2001.  
 XX  
 PF 28-JUL-2000; 2000EP-0116126.  
 XX  
 PR 29-JUL-1999; 99JP-0248036.  
 XX  
 PR 27-AUG-1999; 99JP-0300253.  
 XX  
 PR 11-JAN-2000; 2000JP-0118776.  
 XX  
 PR 09-MAY-2000; 2000JP-0183767.  
 XX  
 PR 09-JUN-2000; 2000JP-0241899.  
 XX  
 PA (HELI-) HELIX RES INST.  
 XX  
 PI Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;  
 PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;  
 XX  
 WP; 2001-318749/34.  
 XX  
 PR Primer sets for synthesizing polynucleotides, particularly the 5602  
 PR full-length cDNAs defined in the specification, and for the detection  
 PR and/or diagnosis of the abnormality of the proteins encoded by the  
 PR full-length cDNAs -  
 XX  
 PS Claim 1; SEQ ID 4364; 2537pp + CD ROM; English.  
 XX  
 CC The present invention describes primer sets for synthesising 5602  
 CC full-length cDNAs defined in the specification. Where a primer set  
 CC comprises: (a) an oligo-dT primer and an oligonucleotide complementary  
 CC to the complementary strand of a polynucleotide which comprises one of  
 CC the 5602 nucleotide sequences defined in the specification, where the  
 CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination  
 CC of an oligonucleotide comprising a sequence complementary to the  
 CC complementary strand of a polynucleotide which comprises a 5'-end  
 CC sequence and an oligonucleotide comprising a sequence complementary to a  
 CC polynucleotide which comprises a 3'-end sequence, where the  
 CC oligonucleotide comprises at least 15 nucleotides and the combination of  
 CC the 5'-end sequence/3'-end sequence is selected from those defined in  
 CC the specification. The primer sets can be used in antisense therapy and  
 CC in gene therapy. The primers are useful for synthesising polynucleotides  
 CC particularly full-length cDNAs. The primers are also useful for the  
 CC detection and/or diagnosis of the abnormality of the proteins encoded by  
 CC the full-length cDNAs. The primers allow obtaining of the full-length  
 CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and  
 CC AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to  
 CC AAB95993 represent human amino acid sequences; and AAH13629 to AAH13632  
 CC represent oligonucleotides, all of which are used in the exemplification  
 CC of the present invention.  
 XX  
 SQ Sequence 807 BP; 240 A; 120 C; 167 G; 274 T; 6 other;

Query Match	90.0%	Score 14.4	DB 22	length 807
Best Local Similarity	93.8%	Pred. No. 4	1e+02	
Matches 15, Conservative	0	Mismatches 1	Indels 0	Gaps 0
0y	1	ccttcaccacacccc	16	



Tue Jul 30 09:10:29 2002

us-09-530-663b-15.rng

Page 13

**THIS PAGE BLANK (USPTO)**

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 29, 2002, 21:29:04 ; Search time 2542.47 Seconds  
(Without alignments)  
84.938 Million cell updates/sec

Title: US-09-530-663b-15  
Perfect score: 16  
Sequence: 1 ccttcaccaccccc 16

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 13736207 seqs, 6748477542 residues  
Total number of hits satisfying chosen parameters: 27472414

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : EST.\*  
1: em\_estdb:\*  
2: em\_esthum:\*  
3: em\_estlin:\*  
4: em\_estnu:\*  
5: em\_estov:\*  
6: em\_estpl:\*  
7: em\_estro:\*  
8: em\_hlc:\*  
9: gb\_estl:\*  
10: gb\_est2:\*  
11: gb\_hlc:\*  
12: gb\_gss:\*  
13: em\_gss\_hum:\*  
14: em\_gss\_inv:\*  
15: em\_gss\_pln:\*  
16: em\_gss\_vrt:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	16	100.0	218	9	BB430167 BB430167
2	16	100.0	445	9	AI452374 mg89g07.x
3	16	100.0	564	12	A2871533 2M0184N21
4	16	100.0	796	12	BH557406 B0GJ056TR
5	16	100.0	837	10	BG771897 602721709
6	16	100.0	1024	12	A2210049 SP_0153_A
7	15	93.8	190	12	A2299798 RPCI-23-4
8	15	93.8	245	12	A2019845 RPCI-23-2
9	15	93.8	329	9	AA503034 nh8B03.s
10	15	93.8	353	12	B82624 RPCI11-17C1
11	15	93.8	372	10	BF766646 IL2-CS004
12	15	93.8	380	9	AV212743 AV212743
13	15	93.8	408	10	BM106270 509959 MA
14	15	93.8	425	9	AI991910 ws42C04.x
15	15	93.8	507	12	AO880479 HS_5044_A
16	15	93.8	626	10	BE347738 sp02909.Y
17	15	93.8	627	9	AV679827 AV679827

C 18	15	93.8	630	10	BE556267	BE556267 sp99e11.Y
C 19	15	93.8	636	10	BF129318	BF129318 601810938
C 20	15	93.8	662	12	AO794741	AO794741 nbxb0053H
C 21	15	93.8	704	12	AG166290	AG166290 Pan trogl
C 22	15	93.8	840	10	BG395823	BG395823 602458604
C 23	15	93.8	866	12	AQ271654	AQ271654 nbxb0026B
C 24	15	93.8	867	10	BI953359	BI953359 HVSMM001
C 25	15	93.8	901	10	BG167717	BG167717 602342919
C 26	15	93.8	910	10	BE414064	BE414064 SCU005.H0
C 27	15	93.8	935	10	BF788352	BF788352 602114092
C 28	15	93.8	938	10	BE285348	BE285348 601095546
C 29	15	93.8	1045	10	BE966779	BE966779 601661280
C 30	15	93.8	1079	10	BG338822	BG338822 602436605
C 31	14.4	90.0	128	9	AV415808	AV415808 AV415808
C 32	14.4	90.0	159	9	BB593966	BB593966 BB593966
C 33	14.4	90.0	166	10	BG964367	BG964367 602831984
C 34	14.4	90.0	167	9	AV313903	AV313903 AV313903
C 35	14.4	90.0	179	9	AV029522	AV029522 AV029522
C 36	14.4	90.0	185	10	BI202286	BI202286 t2H05f.f
C 37	14.4	90.0	186	10	C88801	C88801 C88801 Mous
C 38	14.4	90.0	193	9	AV092502	AV092502 AV092502
C 39	14.4	90.0	197	10	BG981061	BG981061 MR3-CN014
C 40	14.4	90.0	200	9	AU180999	AU180999 AU180999
C 41	14.4	90.0	204	9	BB600337	BB600337 BB600337
C 42	14.4	90.0	210	9	AV354827	AV354827 AV354827
C 43	14.4	90.0	223	9	BA496202	BA496202 BA496202
C 44	14.4	90.0	230	9	BB442511	BB442511 BB442511
C 45	14.4	90.0	231	9	BB514048	BB514048 BB514048

## ALIGNMENTS

RESULT 1  
BB430167  
LOCUS BB430167 218 bp mRNA EST 18-JUL-2000  
DEFINITION BB430167 RIKEN full-length enriched, adult male hippocampus Mus musculus cDNA clone C630031019 3' similar to X86406 R.norvegicus mRNA for brevican, GPI-anchored isoform, mRNA sequence.

ACCESSION BB430167  
KEYWORDS BB430167.1 GI:9269894  
SOURCE house mouse.  
ORGANISM Mus musculus

REFERENCE  
AUTHORS Kono, H., Alzawa, K., Akahira, S., Akiyama, J., Arakawa, T., Carninci, I. (bases 1 to 218)  
P., Endo, T., Fukuda, S., Fukunishi, Y., Hara, A., Hayatsu, N., Hirozane, T., Hori, F., Ishii, Y., Ishikawa, J., Ishikawa, T., Itoh, M., Izawa, M., Kadota, K., Kagawa, I., Kai, C., Kawai, J., Kikuchi, N., Kiyosawa, H., Kojima, Y., Kondo, S., Koya, S., Kurihara, C., Kusakabe, M., Matsuyama, T., Miki, R., Mizuno, T., Nakamura, M., Oda, H., Okazaki, Y., Ono, T., Owa, C., Saito, H., Sakai, C., Sato, K., Shibata, K., Shibata, Y., Shigemoto, Y., Shinagawa, A., Shiraki, T., Sogabe, Y., Sugahara, Y., Suzuki, H., Suzuki, H., Tagawa, A., Takahashi, F., Tomioka, N., Toya, T., Tsunoda, Y., Watabiki, A., Watanabe, S., Yamamura, T., Yamazaki, I., Yano, R., Yasunishi, A., Yokota, T., Yoshida, K., Yoshiki, A., Yoshino, M., Muramatsu, M., and Hayashizaki, Y.

RIKEN Mouse ESTs (Kono, H., et al.)  
Unpublished (2000)  
Contact: Yoshihide Hayashizaki  
Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), Yokohama Institute  
The Institute of Physical and Chemical Research (RIKEN)  
1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan  
Tel: 81-45-503-9222  
Fax: 81-45-503-9216  
Email: genome-resgsc.riken.go.jp,  
URL: http://genome-gsc.riken.go.jp/  
Carninci, P., Nishiyama, Y., Westover, A., Itoh, M., Nagaoka, S., Sasaki, N., Okazaki, Y., Muramatsu, M., and Hayashizaki, Y.

Thermostabilization and thermoactivation of thermolabile enzymes by



/organism="Mus musculus"  
/strain="C57BL/6J"  
/db\_xref="taxon:10090"  
/clone="UUC2M0184N21"  
/clone\_1lb="Mouse 10kb plasmid UUC2M library"  
/sex="Male"  
/lab\_host="E. Coli strain XL10-Gold, T1-resistant, F-"  
/note="Vector: PMD42nv; Purified genomic DNA from M. musculus C57BL/6J (male) was obtained from the Jackson Laboratory Mouse DNA Resource (http://www.jax.org/resources/documents/dnares/). The DNA was hydrodynamically sheared by repeated passage through a 0.005 inch orifice at constant velocity. The sheared DNA was blunt end-repaired with T4 DNA polymerase and T4 polynucleotide kinase. Adaptor oligonucleotides were ligated to the blunt ends in high molar excess. The adapted DNA was purified and size-selected for a 9.5 to 10.5 kb range using preparative agarose gel electrophoresis. Vector DNA was prepared from a derivative of PMD42 (g14732114|gb|AF129072.1), a copy-number inducible derivative of plasmid R1. The vector was ligated with adaptors complementary to the insert adaptors and purified. The sheared, adapted mouse DNA was annealed to adapted vector DNA, and transformed into chemically-competent E. coli XL10-Gold (Stratagene) cells and selected for ampicillin resistance."

BASE COUNT 143 a 123 c 103 g 195 t

ORIGIN

Query Match 100.0%; Score 16; DB 12; Length 564;  
Best Local Similarity 100.0%; Pred. No. 2.4e+03;  
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1 ccttcaccaacccc 16  
|||||

Db 165 CCTTCACCAACCCCC 180

RESULT 4  
BH557406/c 796 bp DNA linear GSS 14-DEC-2001  
LOCUS BOGJ056TR BOGJ Brassica oleracea genomic clone BOGJ056, DNA  
DEFINITION  
ACCESSION BH557406  
VERSION BH557406.1 GI:17809186  
KEYWORDS GSS.  
SOURCE Brassica oleracea.  
ORGANISM Brassica oleracea  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
Rosidae; eurosids II; Brassicales; Brassicaceae; Brassica.  
1 (bases 1 to 796)  
Town,C.D., Van Aken,S., Uterback,T. and Fraser,C.M.  
Whole genome shotgun sequencing of Brassica oleracea  
Unpublished (2001)  
Other GSSs: BOGJ056TF  
Contact: Chris Town  
TIGR Medical Center Drive, Rockville, MD 20850, USA.  
Tel: 301-838-3523  
Fax: 301-838-0208  
Email: cdtown@tigr.org  
DNA is from a doubled haploid provided by Tom Osborn.  
Seq primer: TR  
Class: sheared ends.  
Location/Qualifiers  
1..796  
/organism="Brassica oleracea"  
/strain="T01000DH3"  
/db\_xref="taxon:3712"  
/clone="BOGJ056"  
/clone\_1lb="BOGJ"

FEATURES  
source

/note="Vector: pHO51; site\_1: BstXI; 2-3 kb sheared genomic DNA inserted into pHO51 using BstXI linkers"  
BASE COUNT 199 a 137 c 254 g 206 t

ORIGIN

Query Match 100.0%; Score 16; DB 12; Length 796;  
Best Local Similarity 100.0%; Pred. No. 2.4e+03;  
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1 ccttcaccaacccc 16  
|||||

Db 678 CCTTCACCAACCCCC 663

RESULT 5  
BG771897/c 837 bp mRNA linear EST 15-MAY-2001  
LOCUS BG771897  
DEFINITION 602721709F1 NIH\_MGC\_97 Homo sapiens cDNA clone IMAGE:4838306 5', mRNA sequence.  
ACCESSION BG771897  
VERSION BG771897.1 GI:14082550  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
1 (bases 1 to 837)  
NIH-MGC http://mgc.nci.nih.gov/.  
National Institutes of Health, Mammalian Gene Collection (MGC)  
Unpublished (1999)  
Contact: Robert Strausberg, Ph.D.  
Email: cgapbs-remail.nih.gov  
Tissue Procurement: Miklos Palkovits, M.D., Ph.D.  
cDNA Library Preparation: Michael J. Brownstein (NHGRI), Shiraki  
Toshiyuki and Piero Carninci (RIKEN)  
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
DNA Sequencing by: Incyte Genomics, Inc.  
Clone distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL at:  
http://image.llnl.gov  
Plate: LLM10772 row: b column: 03  
High quality sequence stop: 837.  
Location/Qualifiers  
1..837  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone\_1img="IMAGE:4838306"  
/clone\_1lb="NIH\_MGC\_97"  
/lab\_host="DH10B"  
/note="Organ: testis; Vector: Bluescript (modified pBluescript KS+); Site\_1: BamHI; Site\_2: SalI-XhoI (gicgag); Oligo-dT primed using primer 5'-TTTTTTTTTTVN-3', size-selected for average insert size 2.2 kb and normalized to R0T 5. This is a primary library enriched for full-length clones and constructed using the Cap-trapper method (Carninci, in preparation). Library constructed by M. Brownstein (NIH/NHGRI, National Institutes of Health). Note: this is a NIH\_MGC Library."

BASE COUNT 226 a 179 c 249 g 183 t

ORIGIN

Query Match 100.0%; Score 16; DB 10; Length 837;  
Best Local Similarity 100.0%; Pred. No. 2.4e+03;  
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1 ccttcaccaacccc 16  
|||||

Db 663 CCTTCACCAACCCCC 648

RESULT 6

AZ210049/c 1024 bp DNA linear GSS 31-AUG-2000  
 LOCUS SP.0153-AL.H11.T7A Strongylocentrotus purpuratus, purple sea urchin  
 DEFINITION / sperm genomic BAC library Strongylocentrotus purpuratus genomic  
 clone Plate-153 Col-21 Row-O, DNA sequence.  
 ACCESSION AZ210049  
 VERSION AZ210049.1 GI:8424406  
 KEYWORDS GSS.  
 SOURCE Strongylocentrotus purpuratus.  
 ORGANISM Strongylocentrotus purpuratus.  
 Eukaryota; Metazoa; Echinodermata; Eleutherozoa; Echinozoa;  
 Echinoidae; Euechinoidae; Echinacea; Echinoida;  
 Strongylocentrotidae; Strongylocentrotus.  
 1 (bases 1 to 1024)  
 Cameron,R.A., Mahairas,G., Rast,J.P., Martinez,P., Biondi,T.R.,  
 Smartzell,S., Wallace,J.C., Poustka,A.J., Livingston,B.T., Wray,  
 G.A., Eftensohn,C.A., Lehrach,H., Britten,R.J., Davidson,E.H. and  
 Hood,L.  
 A sea urchin genome project: Sequence scan, virtual map, and  
 additional resources.  
 JOURNAL Proc. Natl. Acad. Sci. U. S. A. 97 (17), 9514-9518 (2000)  
 MEDLINE 20402566  
 COMMENT Contact: Cameron, RA, Davidson, EH, Hood, L  
 Division of Biology 156-29  
 California Institute of Technology  
 Pasadena California 91125, USA  
 Tel: (626) 395-8421  
 Fax: (626) 793-3047  
 Email: hcameron@caltech.edu  
 Plate: 153 row: 0 column: 21  
 Seq primer: T7  
 Class: BAC ends  
 High quality sequence stop: 1024.  
 Location/Qualifiers  
 1..1024  
 /organism="Strongylocentrotus purpuratus"  
 /db\_xref="taxon:7668"  
 /clone\_1fb="Plate=153 Col=21 Row=O"  
 /clone\_1lb="Strongylocentrotus purpuratus, purple sea  
 urchin, sperm genomic BAC library"  
 /note="Organ: sperm; Vector: BACe3.6; BAC Clones in E-Coli  
 DH10B"  
 BASE COUNT 301 a 187 c 212 g 324 t  
 ORIGIN  
 Query Match 100.0%; Score 16; DB 12; Length 1024;  
 Best Local Similarity 100.0%; Pred. No. 2.4e+03;  
 Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 Oy 1 ccttcaccaccccc 16  
 |||||||||||||  
 Db 1020 CCTTCACCAACCCC 1005  
 RESULT 7  
 AZ299798/c 190 bp DNA linear GSS 27-JUL-2000  
 LOCUS RPCI-23-439C15.TV RPCI-23 Mus musculus genomic clone RPCI-23-439C15  
 DEFINITION / DNA sequence.  
 ACCESSION AZ299798  
 VERSION AZ299798.1 GI:9541583  
 KEYWORDS GSS.  
 SOURCE house mouse.  
 ORGANISM Mus musculus.  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 1 (bases 1 to 190)  
 Zhao,S., Nierman,W., Feldblum,T., Malek,J., Shatsman,S., Akinret,  
 B., Levins,M., McGann,S., Tsegaye,G., Geer,K., Krol,M., de Jong,P.  
 and Fraser,C.M.  
 Mouse BAC End Sequences from Library RPCI-23  
 Unpublished (1999)  
 TITLE Mouse BAC End Sequences from Library RPCI-23  
 JOURNAL  
 COMMENT

COMMENT Other\_GSSs: RPCI-23-439C15.TV  
 Contact: Shaying Zhao  
 Department of Eukaryotic Genomics  
 The Institute for Genomic Research  
 9712 Medical Center Dr., Rockville, MD 20850, USA  
 Tel: 301 838 0200  
 Fax: 301 838 0208  
 Email: szhao@tigr.org  
 Clones are derived from the mouse BAC library RPCI-23. For BAC  
 library availability, please contact Pieter de Jong  
 (pieter@dejong.med.buhalo.edu). Clones may be purchased from  
 BACPAC Resources (<http://bacpac.med.buhalo.edu/orderingframe.htm>)  
 or from Resea ch Genetics ([http://www.rtg.org/tb/pac\\_ends/mouse/pac\\_end\\_intro.html](http://www.rtg.org/tb/pac_ends/mouse/pac_end_intro.html))  
 Plate: 439 row: C column: 15  
 Seq primer: T7  
 Class: BAC ends.  
 Location/Qualifiers  
 1..190  
 /organism="Mus musculus"  
 /strain="C57BL/6J"  
 /db\_xref="taxon:10090"  
 /clone="RPCI-23-439C15"  
 /clone\_1lb="RPCI-23"  
 /sex="Female"  
 /lab\_host="DH10B"  
 /note="Organ: Kidney/Brain; Vector: pBACe3.6; Site: 1:  
 EcoRI; Site 2: EcoRI; Female C57BL/6J mouse kidney and/or  
 brain genomic DNA was isolated and partially digested  
 with a combination of EcoRI and EcoRI Methylase. Size  
 selected DNA was cloned into the pBACe3.6 vector at the  
 EcoRI sites. The ligation products were transformed into  
 DH10B electrocompetent cells (BRL Life Technologies)."  
 BASE COUNT 35 a 35 c 36 g 84 t  
 ORIGIN  
 Query Match 93.8%; Score 15; DB 12; Length 190;  
 Best Local Similarity 100.0%; Pred. No. 6.4e+03;  
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 Oy 1 ccttcaccaccccc 15  
 |||||||||||||  
 Db 109 CCTTCACCAACCCC 95  
 RESULT 8  
 AZ019845/c 246 bp DNA linear GSS 25-FEB-2000  
 LOCUS RPCI-23-294J24.TJ RPCI-23 Mus musculus genomic clone RPCI-23-294J24  
 DEFINITION / DNA sequence.  
 ACCESSION AZ019845  
 VERSION AZ019845.1 GI:7095229  
 KEYWORDS GSS.  
 SOURCE house mouse.  
 ORGANISM Mus musculus.  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 1 (bases 1 to 246)  
 Zhao,S., Nierman,W., Feldblum,T., Malek,J., Shatsman,S., Akinret,  
 B., Levins,M., McGann,S., Tsegaye,G., Geer,K., Krol,M., de Jong,P.  
 and Fraser,C.M.  
 Mouse BAC End Sequences from Library RPCI-23  
 Unpublished (1999)  
 TITLE Mouse BAC End Sequences from Library RPCI-23  
 JOURNAL  
 COMMENT Contact: Shaying Zhao  
 Department of Eukaryotic Genomics  
 The Institute for Genomic Research  
 9712 Medical Center Dr., Rockville, MD 20850, USA  
 Tel: 301 838 0200  
 Fax: 301 838 0208  
 Email: szhao@tigr.org  
 Clones are derived from the mouse BAC library RPCI-23. For BAC  
 library availability, please contact Pieter de Jong

(pieterdejong.med.buffalo.edu). Clones may be purchased from BACPAC Resources (<http://bacpac.med.buffalo.edu/orderingframe.htm>) or from Resea ch Genetics (<http://resgen.com>). BAC end page: [http://www.tlgr.org/tldb/bac-ends/mouse/bac\\_end\\_intro.html](http://www.tlgr.org/tldb/bac-ends/mouse/bac_end_intro.html)  
 Plate: 294 row: 5 column: 24  
 Seq primer: SP6  
 Class: BAC ends.

#### FEATURES

source 1. 246  
 Location/Qualifiers  
 /organism="Mus musculus"  
 /strain="C57BL/6J"  
 /db\_xref="taxon:10090"  
 /clone="RPCI-23-294J24"  
 /clone\_1lb="RPCI-23"  
 /sex="Female"  
 /lab\_host="DH10B"  
 /note="Organ: Kidney/Brain; Vector: pBACe3.6; Site\_1: EcoRI; Site\_2: EcoRI; Female C57BL/6J mouse kidney and/or brain genomic DNA was isolated and partially digested with a combination of EcoRI and EcoRI Methylase. Size selected DNA was cloned into the pBACe3.6 vector at the EcoRI sites. The ligation products were transformed into DH10B electrocompetent cells (BRL Life Technologies)."  
 BASE COUNT 49 a 41 c 48 g 108 t  
 ORIGIN

Query Match 93.8%; Score 15; DB 12; Length 246;  
 Best Local Similarity 100.0%; Pred. No. 6.5e+03;  
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 ccttcaccaacccc 15  
 |||||||||||||  
 Db 108 CCTTCACCAACCCC 94

RESULT 9  
 AA503034 329 bp mRNA linear EST 19-AUG-1997  
 LOCUS nhsb03.s1 NCI-CGAP Pr8 Homo sapiens CDNA clone IMAGE:956525  
 DEFINITION similar to gb:DI0667 MYOSIN HEAVY CHAIN, SMOOTH MUSCLE ISOFORM (HUMAN); contains element PK7 repetitive element; mRNA sequence.

ACCESSION AA503034 GI:2238001  
 VERSION AA503034  
 KEYWORDS EST.  
 SOURCE human.  
 ORGANISM Homo sapiens

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
 1 (bases 1 to 329)  
 NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.  
 National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
 Tumor Gene Index

JOURNAL Unpublished (1997)  
 COMMENT Contact: Robert Strausberg, Ph.D.  
 Email: [cgabp-sr@mail.nih.gov](mailto:cgabp-sr@mail.nih.gov)  
 Tissue Procurement: David G. Bostwick, M.D., Rodrigo F. Chuquai,  
 M.D., Michel R. Emmert-Buck, M.D., Ph.D.  
 CDNA Library Preparation: David B. Krizman, Ph.D.  
 DNA Library Arrayed by: Greg Lennon, Ph.D.  
 DNA Sequencing by: Washington University Genome Sequencing Center  
 Clone distribution: NCI-CGAP clone distribution information can be  
 found through the I.M.A.G.E. Consortium/LLNL at:  
[www-bio.llnl.gov/bbrp/image/image.html](http://www-bio.llnl.gov/bbrp/image/image.html)  
 Seq primer: 40m13 fwd. RT from Amersham.

#### FEATURES

source 1. 329  
 Location/Qualifiers  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:956525"  
 /clone\_1lb="NCI-CGAP\_Pr8"  
 /sex="male"  
 /tissue\_type="prostate"

/lab\_host="DH10B"  
 /note="Vector: PAMP10; mRNA made from invasive prostate tumor. CDNA made by oligo-dT priming. Non-directionally cloned. Size-selected on agarose gel, average insert size 600 bp."  
 BASE COUNT 53 a 82 c 73 g 121 t  
 ORIGIN

Query Match 93.8%; Score 15; DB 9; Length 329;  
 Best Local Similarity 100.0%; Pred. No. 6.5e+03;  
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2 ctttcaccaacccc 16  
 |||||||||||||  
 Db 173 CTTTCACCAACCCC 159

RESULT 10  
 B82624/c 353 bp DNA linear GSS 09-APR-1999  
 LOCUS RPC111-17C17.TP RPCI-11 Homo sapiens genomic clone RPCI-11-17C17,  
 DEFINITION DNA sequence.

ACCESSION B82624 GI:2869647  
 VERSION B82624  
 KEYWORDS GSS.  
 SOURCE human.  
 ORGANISM Homo sapiens

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
 1 (bases 1 to 353)  
 Adams, M.D., Rounsley, S.D., Zhao, S., Field, C.E., Bass, S., Linher, K.,  
 Golden, K., Berry, K., Granger, D., Suh, E., Wible, C., de Jong, P. and  
 Venter, J.C.  
 Use of BAC End Sequences for Sequence-Ready Map Building (1998)  
 Unpublished (1998)  
 Other\_GSSs: RPC111-17C17.TV  
 CONTACT: Mark Adams  
 Department of Eukaryotic Genomics  
 The Institute for Genomic Research  
 9712 Medical Center Dr., Rockville, MD 20850, USA  
 Tel: 301 838 0200  
 Fax: 301 838 0208  
 Email: [mdadams@tlgr.org](mailto:mdadams@tlgr.org)

Clones are derived from the human BAC library RPCI-11. For BAC library availability, please contact Pieter de Jong ([pieterdejong.med.buffalo.edu](mailto:pieterdejong.med.buffalo.edu)). Clones may be purchased from BACPAC Resources (<http://bacpac.med.buffalo.edu/ordering>) or from Research Genetics (<http://resgen.com>). BAC end search page: [http://www.tlgr.org/tldb/humgen/bac\\_end\\_search/bac\\_end\\_search.html](http://www.tlgr.org/tldb/humgen/bac_end_search/bac_end_search.html)  
 Seq primer: SP6  
 Class: BAC ends.

#### FEATURES

source 1. 353  
 Location/Qualifiers  
 /organism="Homo sapiens"  
 /db\_xref="GDB:7506208"  
 /db\_xref="taxon:9606"  
 /clone="RPCI-11-17C17"  
 /clone\_1lb="RPCI-11"  
 /sex="Male"  
 /cell\_type="Lymphocytes"  
 /note="Vector: pBACe3.6; Site\_1: EcoRI; Site\_2: EcoRI;  
 RPC11 Human Male BAC library"  
 BASE COUNT 102 a 60 c 65 g 126 t  
 ORIGIN

Query Match 93.8%; Score 15; DB 12; Length 353;  
 Best Local Similarity 100.0%; Pred. No. 6.5e+03;  
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 ccttcaccaacccc 15  
 |||||||||||||





QY 2 cttcaccaccccc 16  
 DB 131 CTTTACCACCCCC 145

RESULT 13  
 BM106270 408 bp mRNA linear EST 21-NOV-2001

LOCUS  
 DEFINITION 509955 MARC 3BOV Bos taurus CDNA 5', mRNA sequence.  
 ACCESSION BM106270  
 VERSION BM106270.1 GI:17037340  
 KEYWORDS EST.  
 SOURCE cow.  
 ORGANISM Bos taurus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;  
 Bovidae; Bovinae; Bos.

REFERENCE  
 1 (bases 1 to 408)  
 Smith,T.P.L., Grosse,W.M., Freking,B.A., Roberts,A.J., Stone,R.T.,  
 Casas,E., Wray,J.E., White,J., Cho,J., Fahrenkrug,S.C., Bennett,  
 G.L., Heaton,M.P., Laegreid,W.W., Kohrer,G.A., Chitko-McKown,C.G.,  
 Potee,G., Holt,I., Karameycheva,S., Liang,F., Quackenbush,J. and  
 Keeler,J.W.  
 Sequence evaluation of four pooled-tissue normalized bovine CDNA  
 libraries and construction of a gene index for cattle  
 Genome Res. 11 (4), 626-630 (2001)

TITLE  
 JOURNAL 21180013  
 MEDLINE  
 COMMENT  
 Contact: Smith TPL  
 USDA, ARS, US Meat Animal Research Center  
 PO Box 166, Clay Center, NE 68933-0166, USA  
 Tel: 402 762 4366  
 Fax: 402 762 4390  
 Email: smithemall.marc.usda.gov  
 Single pass sequencing. Bases called and alt-trimmed with phred  
 v0.980904.e. Vector identified by cross-match with the -m1nscore 18  
 and -minmatch 12 options.  
 PCR primers  
 FORWARD: AGGAACAGCATATGACCAT  
 BACKWARD: GTTTCCTCAGTACGACG  
 Plate: 103 Row: 0 Column: 12  
 Seq primer: AATTAGTGACATATAG.  
 Location/Qualifiers  
 1..408  
 /organism="Bos taurus"  
 /db\_xref="taxon:9913"  
 /clone\_lib="MARC 3BOV"  
 /tissue\_type="pooled"  
 /lab\_host="DH10B"  
 /note="Vector: pCMV SPORT6; Site.1: XbaI; Site.2: XhoI;  
 library made from pooled tissue from marrow, alveolar  
 macrophage, ovary, fetal semitendinosus muscle, and fetal  
 longissimus muscle."

BASE COUNT 94 a 94 c 102 g 118 t  
 ORIGIN

Query Match 93.8%; Score 15; DB 10; Length 408;  
 Best Local Similarity 100.0%; Pred. No. 6.5e+03;  
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 cttcaccaccccc 15  
 DB 226 CTTTACCACCCCC 240

RESULT 14  
 A1991910 425 bp mRNA linear EST 08-MAR-2000  
 LOCUS A1991910/c  
 DEFINITION w642c04.x1 NCI-CGAP Brn25 Homo sapiens CDNA clone IMAGE:249846 3'  
 Similar to contains TARI.tl TARI Repetitive element ;, mRNA  
 Sequence.  
 ACCESSION A1991910

VERSION A1991910.1 GI:5838815  
 KEYWORDS EST.  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE  
 1 (bases 1 to 425)  
 NCI/NINDS-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.  
 National Cancer Institute / National Institute of Neurological  
 Disorders and Stroke, Brain Tumor Genome Anatomy Project  
 (CGAP/BTCAP), Tumor Gene Index  
 Unpublished (1998)  
 JOURNAL Contact: Robert Strausberg, Ph.D.  
 COMMENT Email: cgaps-email.nih.gov  
 Tissue Procurement: David N. Louis, M.D., Myrna R. Rosenfeld M.D.,  
 Ph.D.  
 cDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima  
 Bonaldo, Ph.D.  
 cDNA Library Arrayed by: Greg Lennon, Ph.D.  
 DNA Sequencing by: Washington University Genome Sequencing Center  
 Clone distribution: NCI-CGAP clone distribution information can be  
 found through the I.M.A.G.E. Consortium/LLNL at:  
 www-bio.llnl.gov/dbp/Image/Image.html  
 Insert Length: 822 Std Error: 0.00  
 Seq primer: -40UP from Gibco  
 High quality sequence stop: 416.  
 Location/Qualifiers  
 1..425  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone\_image="IMAGE:249846"  
 /clone\_lib="NCI-CGAP\_Brn25"  
 /tissue\_type="anaplastic oligodendroglioma"  
 /lab\_host="DH10B"  
 /note="Organ: brain; Vector: p773D-Pec (Pharmacia) with a  
 modified polylinker; Site.1: Not I; Site.2: Eco RI; 1st  
 strand cDNA was primed with a Not I - oligo(dT) primer [5'  
 TGTTCACATCTGAAAGTGAGCGGCGCATGAGTGTGTGTGTGTGTGT  
 T 3']; double-stranded cDNA was ligated to Eco RI  
 adaptors (Pharmacia), digested with Not I and cloned into  
 the Not I and Eco RI sites of the modified p773 vector.  
 Library is normalized, and was constructed by Bento  
 Soares and M.Fatima Bonaldo."

BASE COUNT 83 a 151 c 125 g 63 t 3 others  
 ORIGIN

Query Match 93.8%; Score 15; DB 9; Length 425;  
 Best Local Similarity 100.0%; Pred. No. 6.5e+03;  
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 cttcaccaccccc 15  
 DB 358 CTTTACCACCCCC 344

RESULT 15  
 A0880479 507 bp DNA linear GSS 09-NOV-1999  
 LOCUS HS\_5044\_A2-G02-SP6E RPCI-11 Human Male BAC Library Homo sapiens  
 DEFINITION genomic clone Plate=8812 Col=4 Row=M, DNA sequence.  
 ACCESSION A0880479  
 VERSION A0880479.1 GI:6311946  
 KEYWORDS GSS.  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE  
 1 (bases 1 to 507)  
 Mahairis,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T.,  
 Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and  
 Hood,L.  
 Sequence-tagged connectors: A sequence approach to mapping and

JOURNAL  
MEDLINE  
COMMENT

scanning the human genome  
Proc. Natl. Acad. Sci. U. S. A. 96 (17), 9739-9744 (1999)  
99380589  
Contact: Mahairas GG, Wallace JC, Hood L  
High Throughput Sequencing Center  
University of Washington  
401 Queen Anne Avenue North, Seattle, WA 98109, USA  
Tel: (206) 616-3618  
Fax: (206) 616-3887  
Email: jwallace@u.washington.edu  
Clones are derived from the human BAC library RPCI-11. For BAC  
library availability, please contact Pieter de Jong  
(pieter@dejong.med.bufileo.edu). Clones may be purchased from  
BACPAC Resources ([http://bacpac.med.bufileo.edu/ordering\\_bac.htm](http://bacpac.med.bufileo.edu/ordering_bac.htm))  
or from Research Genetics (<http://www.htsc.washington.edu>). BAC end Web Server:  
<http://www.htsc.washington.edu>  
Plate: 8812 Row: M Column: 4  
Seq primer: SP6  
Class: BAC ends  
High quality sequence stop: 507.  
Location/Qualifiers  
1.507  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="Plate=8812 Col=4 Row=M"  
/clone\_lib="RPCI-11 Human Male BAC Library"  
/sex="male"  
/note="Vector: pBACe3.6; Site\_1: EcoRI; Site\_2: EcoRI;  
Male blood DNA was isolated from one randomly chosen donor  
and partially digested with a combination of EcoRI and  
EcoRI Methylase. Size selected DNA was cloned into the  
pBACe3.6 vector at EcoRI sites"

BASE COUNT 141 a 117 c 107 g 135 t 7 others  
ORIGIN

Query Match 93.8%; Score 15; DB 12; Length 507;  
Best Local Similarity 100.0%; Pred. No. 6.5e+03;  
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 2 cttccaccaccccc 16  
|||||  
Db 385 CTTCCACCAACCCCC 371

Search completed: July 29, 2002, 23:22:40  
Job time: 6816 sec

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 29, 2002, 22:36:59 ; Search time 65.09 seconds  
(without alignments)  
60.380 Million cell updates/sec

Title: US-09-530-663B-15  
Perfect score: 16  
Sequence: 1 ccttcaccaccccc 16

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 383533 seqs, 122816752 residues

Total number of hits satisfying chosen parameters: 767066

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 10%  
Listing first 45 summaries

Database : Issued\_Patents\_NA: \*  
1: /cgn2\_6/ptodata/2/ina/5A.COMB.seq: \*  
2: /cgn2\_6/ptodata/2/ina/5B.COMB.seq: \*  
3: /cgn2\_6/ptodata/2/ina/6A.COMB.seq: \*  
4: /cgn2\_6/ptodata/2/ina/6B.COMB.seq: \*  
5: /cgn2\_6/ptodata/2/ina/PCRTUS.COMB.seq: \*  
6: /cgn2\_6/ptodata/2/ina/Backfile1.seq: \*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	14.4	90.0	1100	3	US-08-978-741-16 Sequence 16, Appl
2	14.4	90.0	1100	4	US-09-333-729A-16 Sequence 16, Appl
3	14.4	90.0	1300	3	US-08-978-741-4 Sequence 4, Appl
4	14.4	90.0	1300	4	US-09-333-729A-6 Sequence 6, Appl
5	14.4	90.0	1514	3	US-08-978-741-1 Sequence 1, Appl
6	14.4	90.0	1514	4	US-09-333-729A-2 Sequence 2, Appl
7	14.4	90.0	5009	3	US-08-978-741-7 Sequence 7, Appl
8	14.4	90.0	5009	4	US-09-333-729A-8 Sequence 8, Appl
9	14.4	90.0	11284	3	US-08-978-741-5 Sequence 5, Appl
10	14.4	87.5	79	4	US-08-463-903-78 Sequence 78, Appl
11	14.4	87.5	95	4	US-07-935-695-78 Sequence 78, Appl
12	14.4	87.5	95	3	US-08-463-903-79 Sequence 79, Appl
13	14.4	87.5	95	4	US-07-935-695-79 Sequence 79, Appl
14	13.4	83.8	206	2	US-08-485-657A-4 Sequence 4, Appl
15	13.4	83.8	208	5	PCT-US95-02303-4 Sequence 4, Appl
16	13.4	83.8	559	4	US-08-975-762-4 Sequence 4, Appl
17	13.4	83.8	559	4	US-08-821-324-4 Sequence 4, Appl
18	13.4	83.8	559	4	US-09-295-028-4 Sequence 4, Appl
19	13.4	83.8	559	4	US-09-106-582-4 Sequence 4, Appl
20	13.4	83.8	1330	2	US-08-933-750C-80 Sequence 80, Appl
21	13.4	83.8	1330	3	US-09-234-613-80 Sequence 80, Appl
22	13.4	83.8	1656	1	US-08-324-465-2 Sequence 2, Appl
23	13.4	83.8	1656	2	US-08-465-981-2 Sequence 2, Appl
24	13.4	83.8	1656	5	PCT-US93-11915-2 Sequence 5, Appl
25	13.4	83.8	1725	1	US-08-324-465-5 Sequence 5, Appl
26	13.4	83.8	1725	2	US-08-465-981-5 Sequence 5, Appl
27	13.4	83.8	1725	5	PCT-US93-11915-5 Sequence 5, Appl

28	13.4	83.8	1821	1	US-07-803-622E-3 Sequence 3, Appl
c 29	13.4	83.8	1919	4	US-08-975-762-40 Sequence 40, Appl
c 30	13.4	83.8	1919	4	US-09-295-028-40 Sequence 40, Appl
c 31	13.4	83.8	1919	4	US-09-106-582-40 Sequence 40, Appl
c 32	13.4	83.8	2265	2	US-08-940-332-1 Sequence 1, Appl
c 33	13.4	83.8	6263	2	US-08-781-802-3 Sequence 1, Appl
34	13.4	83.8	6263	4	US-08-694-078-3 Sequence 3, Appl
35	13.4	83.8	8779	2	US-08-750-703-4 Sequence 4, Appl
36	13.4	83.8	12752	2	US-08-459-146-1 Sequence 1, Appl
37	13.4	83.8	12752	2	US-08-459-065-1 Sequence 1, Appl
38	13.4	81.2	1814	2	US-08-483-151-1 Sequence 1, Appl
39	13.4	81.2	1814	5	PCT-US96-06427-1 Sequence 1, Appl
40	13.4	81.2	3116	4	US-09-362-831-10 Sequence 10, Appl
41	13.4	81.2	3624	1	US-07-951-715A-6 Sequence 6, Appl
42	13.4	81.2	3624	2	US-08-459-448A-6 Sequence 6, Appl
43	13.4	81.2	3624	3	US-08-459-505A-6 Sequence 6, Appl
44	13.4	81.2	3624	3	US-08-459-504B-6 Sequence 6, Appl
45	13.4	81.2	3624	3	US-08-459-444-6 Sequence 6, Appl

## ALIGNMENTS

```
RESULT 1
US-08-978-741-16
: Sequence 16, Application US/08978741
: Patent No. 6100076
:
: GENERAL INFORMATION:
: APPLICANT: Yang Wang, Michael W. Spellman
: TITLE OF INVENTION: O-Fucosyltransferase
: NUMBER OF SEQUENCES: 17
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Genentech, Inc.
: STREET: 1 DNA Way
: CITY: South San Francisco
: STATE: California
: COUNTRY: USA
: ZIP: 94080
: COMPUTER READABLE FORM:
: MEDIUM TYPE: 3.5 inch, 1.44 Mb floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Winpatin (Genentech)
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/978, 741
: FILING DATE: 26-NO. 6100076-1997
: CLASSIFICATION: 435
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: 08/792498
: FILING DATE: 31
: ATTORNEY/AGENT INFORMATION:
: NAME: Svoboda, Craig G.
: REGISTRATION NUMBER: 39,044
: REFERENCE/DOCKET NUMBER: P1041P1
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 650/225-1489
: TELEFAX: 650/952-9861
: INFORMATION FOR SEQ ID NO: 16:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 1100 base pairs
: TYPE: Nucleic Acid
: STRANDEDNESS: Single
: TOPOLOGY: Linear
:
: US-08-978-741-16
:
: Query Match 90.0%; Score 14.4; DB 3; Length 1100;
: Best Local Similarity 93.8%; Pred. No. 51;
: Matches 15; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
:
: QY 1 ccttcaccaccccc 16
: DB 166 CCTTCACCACTCC 181
```

RESULT 2  
US-09-333-729A-16  
; Sequence 16, Application US/09333729A  
; Patent No. 6270987  
; GENERAL INFORMATION:  
; APPLICANT: Wang, Yang  
; APPLICANT: Spellman, Michael W.  
; TITLE OF INVENTION: O-Fucosyltransferase  
; FILE REFERENCE: P1041PDI-Substitute  
; CURRENT APPLICATION NUMBER: US/09/333,729A  
; CURRENT FILING DATE: 1999-06-15  
; PRIOR APPLICATION NUMBER: US 08/798,741  
; PRIOR FILING DATE: 1997-11-26  
; NUMBER OF SEQ ID NOS: 21  
; SEQ ID NO 16  
; LENGTH: 1100  
; TYPE: DNA  
; ORGANISM: Homo Sapien  
US-09-333-729A-16

Query Match 90.0%; Score 14.4; DB 4; Length 1100;  
Best Local Similarity 93.8%; Pred. No. 51;  
Matches 15; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 ccttcaccaccccc 16  
|||||  
DB 166 ccttcaccacctcc 181

RESULT 3  
US-08-978-741-4  
; Sequence 4, Application US/08978741  
; Patent No. 6100076  
; GENERAL INFORMATION:  
; APPLICANT: Yang Wang, Michael W. Spellman  
; TITLE OF INVENTION: O-Fucosyltransferase  
; NUMBER OF SEQUENCES: 17  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Genentech, Inc.  
; STREET: 1 DNA Way  
; CITY: South San Francisco  
; STATE: California  
; COUNTRY: USA  
; ZIP: 94080  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: 3.5 inch, 1.44 Mb floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Winpatin (Genentech)  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/978,741  
; FILING DATE: 26-No. 6100076-1997  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 08/792498  
; FILING DATE: 31  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Svoboda, Craig G.  
; REGISTRATION NUMBER: 39,044  
; REFERENCE/DOCKET NUMBER: P1041P1  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 650/225-1489  
; TELEFAX: 650/952-9881  
; INFORMATION FOR SEQ ID NO: 4:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1300 base pairs  
; TYPE: Nucleic Acid  
; STRANDEDNESS: Single  
; TOPOLOGY: Linear  
US-08-978-741-4

Query Match 90.0%; Score 14.4; DB 3; Length 1300;  
Best Local Similarity 93.8%; Pred. No. 52;  
Matches 15; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 ccttcaccaccccc 16  
|||||  
DB 301 ccttcaccacacccctcc 316

RESULT 4  
US-09-333-729A-6  
; Sequence 6, Application US/09333729A  
; Patent No. 6270987  
; GENERAL INFORMATION:  
; APPLICANT: Wang, Yang  
; APPLICANT: Spellman, Michael W.  
; TITLE OF INVENTION: O-Fucosyltransferase  
; FILE REFERENCE: P1041PDI-Substitute  
; CURRENT APPLICATION NUMBER: US/09/333,729A  
; CURRENT FILING DATE: 1999-06-15  
; PRIOR APPLICATION NUMBER: US 08/798,741  
; PRIOR FILING DATE: 1997-11-26  
; NUMBER OF SEQ ID NOS: 21  
; SEQ ID NO 6  
; LENGTH: 1300  
; TYPE: DNA  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Plasmid insert.  
US-09-333-729A-6

Query Match 90.0%; Score 14.4; DB 4; Length 1300;  
Best Local Similarity 93.8%; Pred. No. 52;  
Matches 15; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 ccttcaccaccccc 16  
|||||  
DB 301 ccttcaccacacccctcc 316

RESULT 5  
US-08-978-741-1  
; Sequence 1, Application US/08978741  
; Patent No. 6100076  
; GENERAL INFORMATION:  
; APPLICANT: Yang Wang, Michael W. Spellman  
; TITLE OF INVENTION: O-Fucosyltransferase  
; NUMBER OF SEQUENCES: 17  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Genentech, Inc.  
; STREET: 1 DNA Way  
; CITY: South San Francisco  
; STATE: California  
; COUNTRY: USA  
; ZIP: 94080  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: 3.5 inch, 1.44 Mb floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Winpatin (Genentech)  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/978,741  
; FILING DATE: 26-No. 6100076-1997  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 08/792498  
; FILING DATE: 31  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Svoboda, Craig G.  
; REGISTRATION NUMBER: 39,044

```

; REFERENCE/DOCKET NUMBER: P1041P1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 650/225-1489
; TELEFAX: 650/952-9881
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1514 base pairs
; TYPE: Nucleic Acid
; STRANDEDNESS: Single
; TOPOLOGY: Linear
US-08-978-741-1

Query Match          90.0%; Score 14.4; DB 3; Length 1514;
Best Local Similarity 93.8%; Pred. No. 53;
Matches 15; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 ccttcaccaacccc 16
   |||||
Db 166 ccttcaccaacccc 181

RESULT 6
US-09-333-729A-2
; Sequence 2, Application US/09333729A
; Patent No. 6270987
; GENERAL INFORMATION:
; APPLICANT: Wang, Yang
; TITLE OF INVENTION: O-Fucosyltransferase
; FILE REFERENCE: P1041P1D1-Substitute
; CURRENT APPLICATION NUMBER: US/09/333,729A
; CURRENT FILING DATE: 1999-06-15
; PRIOR APPLICATION NUMBER: US 08/798,741
; PRIOR FILING DATE: 1997-11-26
; NUMBER OF SEQ ID NOS: 21
; SEQ ID NO 2
; LENGTH: 1514
; TYPE: DNA
; ORGANISM: Homo Sapien
US-09-333-729A-2

Query Match          90.0%; Score 14.4; DB 4; Length 1514;
Best Local Similarity 93.8%; Pred. No. 53;
Matches 15; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 ccttcaccaacccc 16
   |||||
Db 166 ccttcaccaacccc 181

RESULT 7
US-08-978-741-7
; Sequence 7, Application US/08978741
; Patent No. 6100076
; GENERAL INFORMATION:
; APPLICANT: Yang Wang, Michael W. Spellman
; TITLE OF INVENTION: O-Fucosyltransferase
; NUMBER OF SEQUENCES: 17
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genentech, Inc.
; STREET: 1 DNA Way
; CITY: South San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94080
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5 inch, 1.44 Mb floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: WinPatIn (Genentech)
; CURRENT APPLICATION DATA:
```

```

; APPLICATION NUMBER: US/08/978,741
; FILING DATE: 26-No. 6100076-1997
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/792498
; FILING DATE: 31
; ATTORNEY/AGENT INFORMATION:
; NAME: Svoboda, Craig G.
; REGISTRATION NUMBER: 39,044
; REFERENCE/DOCKET NUMBER: P1041P1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 650/225-1489
; TELEFAX: 650/952-9881
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 5009 base pairs
; TYPE: Nucleic Acid
; STRANDEDNESS: Single
; TOPOLOGY: Linear
US-08-978-741-7

Query Match          90.0%; Score 14.4; DB 3; Length 5009;
Best Local Similarity 93.8%; Pred. No. 61;
Matches 15; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 ccttcaccaacccc 16
   |||||
Db 101 ccttcaccaacccc 116

RESULT 8
US-09-333-729A-8
; Sequence 8, Application US/09333729A
; Patent No. 6270987
; GENERAL INFORMATION:
; APPLICANT: Wang, Yang
; TITLE OF INVENTION: O-Fucosyltransferase
; FILE REFERENCE: P1041P1D1-Substitute
; CURRENT APPLICATION NUMBER: US/09/333,729A
; CURRENT FILING DATE: 1999-06-15
; PRIOR APPLICATION NUMBER: US 08/798,741
; PRIOR FILING DATE: 1997-11-26
; NUMBER OF SEQ ID NOS: 21
; SEQ ID NO 8
; LENGTH: 5009
; TYPE: DNA
; ORGANISM: Homo Sapien
US-09-333-729A-8

Query Match          90.0%; Score 14.4; DB 4; Length 5009;
Best Local Similarity 93.8%; Pred. No. 61;
Matches 15; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 ccttcaccaacccc 16
   |||||
Db 101 ccttcaccaacccc 116

RESULT 9
US-08-978-741-5
; Sequence 5, Application US/08978741
; Patent No. 6100076
; GENERAL INFORMATION:
; APPLICANT: Yang Wang, Michael W. Spellman
; TITLE OF INVENTION: O-Fucosyltransferase
; NUMBER OF SEQUENCES: 17
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genentech, Inc.
; STREET: 1 DNA Way
; CITY: South San Francisco
```

STATE: California  
COUNTRY: USA  
ZIP: 94080  
COMPUTER READABLE FORM:  
MEDIUM TYPE: 3.5 inch, 1.44 Mb floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: WinPatIn (Genentech)  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/978,741  
FILING DATE: 26-No. 6100076-1997  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/792498  
FILING DATE: 31  
ATTORNEY/AGENT INFORMATION:  
NAME: Syvoda, Craig G.  
REGISTRATION NUMBER: 39,044  
REFERENCE/DOCKET NUMBER: P1041P1  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 650/225-1489  
TELEFAX: 650/952-9881  
INFORMATION FOR SEQ ID NO: 5:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 11284 base pairs  
TYPE: Nucleic Acid  
STRANDEDNESS: Single  
TOPOLOGY: Linear  
US-08-978-741-5

Query Match 90.0%; Score 14.4; DB 3; Length 11284;  
Best Local Similarity 93.8%; Pred. No. 66;  
Matches 15; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
QY 1 ccttcaccaacccc 16  
|||||  
Db 4401 CTTTCACCAACCTCC 4416

RESULT 10  
US-08-463-903-78/c  
Sequence 78, Application US/08463903  
Patent No. 6071515  
GENERAL INFORMATION:  
APPLICANT: Mezes, Peter S.  
APPLICANT: Richard, Ruth A.  
APPLICANT: Affholter, Joseph A.  
APPLICANT: Kollte, Nicolas J.  
TITLE OF INVENTION: Dimer and Multimer Forms of Single Chain Polypeptides  
FILE REFERENCE: 40224A US  
CURRENT APPLICATION NUMBER: US/08/463,903  
CURRENT FILING DATE: 1995-06-05  
EARLIER APPLICATION NUMBER: US 07/935,695  
EARLIER FILING DATE: 1992-08-21  
NUMBER OF SEQ ID NOS: 102  
SOFTWARE: MS-Word for Windows, Ver. 7.0  
SEQ ID NO 78  
LENGTH: 79  
TYPE: DNA  
ORGANISM: Artificial Sequence  
FEATURE:  
NAME/KEY: DRb-3AH5' primer  
LOCATION: 1..79  
US-08-463-903-78

Query Match 87.5%; Score 14; DB 3; Length 79;  
Best Local Similarity 100.0%; Pred. No. 61;  
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 2 ctttcaccaacccc 15  
|||||

Db 68 CTTTCACCAACCCC 55

RESULT 11  
US-07-935-695-78/c  
Sequence 78, Application US/07935695  
Patent No. 6329507  
GENERAL INFORMATION:  
APPLICANT: Mezes, Peter S.  
APPLICANT: Richard, Ruth A.  
APPLICANT: Affholter, Joseph A.  
APPLICANT: Kollte, Nicolas J.  
TITLE OF INVENTION: Dimer and Multimer Forms of Single Chain Polypeptides  
FILE REFERENCE: 40224A US  
CURRENT APPLICATION NUMBER: US/07/935,695  
CURRENT FILING DATE: 1992-08-21  
PRIOR APPLICATION NUMBER: US 08/463,903  
PRIOR FILING DATE: 1995-06-05  
NUMBER OF SEQ ID NOS: 102  
SOFTWARE: MS-Word for Windows, Ver. 7.0  
SEQ ID NO 78  
LENGTH: 79  
TYPE: DNA  
ORGANISM: Artificial Sequence  
FEATURE:  
NAME/KEY: DRb-3AH5' primer  
LOCATION: 1..79  
OTHER INFORMATION: :  
US-07-935-695-78

Query Match 87.5%; Score 14; DB 4; Length 79;  
Best Local Similarity 100.0%; Pred. No. 61;  
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 2 ctttcaccaacccc 15  
|||||  
Db 68 CTTTCACCAACCCC 55

RESULT 12  
US-08-463-903-79  
Sequence 79, Application US/08463903  
Patent No. 6071515  
GENERAL INFORMATION:  
APPLICANT: Mezes, Peter S.  
APPLICANT: Richard, Ruth A.  
APPLICANT: Affholter, Joseph A.  
APPLICANT: Kollte, Nicolas J.  
TITLE OF INVENTION: Dimer and Multimer Forms of Single Chain Polypeptides  
FILE REFERENCE: 40224A US  
CURRENT APPLICATION NUMBER: US/08/463,903  
CURRENT FILING DATE: 1995-06-05  
EARLIER APPLICATION NUMBER: US 07/935,695  
EARLIER FILING DATE: 1992-08-21  
NUMBER OF SEQ ID NOS: 102  
SOFTWARE: MS-Word for Windows, Ver. 7.0  
SEQ ID NO 79  
LENGTH: 95  
TYPE: DNA  
ORGANISM: Artificial Sequence  
FEATURE:  
NAME/KEY: DRb-3AHV3' primer  
LOCATION: 1..95  
US-08-463-903-79

Query Match 87.5%; Score 14; DB 3; Length 95;  
Best Local Similarity 100.0%; Pred. No. 63;  
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 2 ctttcaccaacccc 15  
|||||







GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 29, 2002, 23:55:06 ; Search time 1921.77 Seconds  
(without alignments)  
119.781 Million cell updates/sec

Title: US-09-530-663B-17  
Perfect score: 11  
Sequence: 1 tctcaccacc 11

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 1797656 segs, 10463268293 residues  
Total number of hits satisfying chosen parameters: 3595312

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : GenEmbl:  
1: gb\_da:\*  
2: gb\_htg:\*  
3: gb\_in:\*  
4: gb\_cm:\*  
5: gb\_ov:\*  
6: gb\_pat:\*  
7: gb\_ph:\*  
8: gb\_pl:\*  
9: gb\_pr:\*  
10: gb\_ro:\*  
11: gb\_sts:\*  
12: gb\_sy:\*  
13: gb\_un:\*  
14: gb\_vl:\*  
15: em\_ba:\*  
16: em\_fun:\*  
17: em\_hum:\*  
18: em\_in:\*  
19: em\_mu:\*  
20: em\_on:\*  
21: em\_or:\*  
22: em\_ov:\*  
23: em\_pat:\*  
24: em\_ph:\*  
25: em\_pl:\*  
26: em\_ro:\*  
27: em\_sts:\*  
28: em\_un:\*  
29: em\_vl:\*  
30: em\_htg\_hum:\*  
31: em\_htg\_inv:\*  
32: em\_htg\_other:\*  
33: em\_htgo\_inv:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result Query Match Length DB ID Description  
No. Score

1	11	100.0	12	6	I24588	I24588 Sequence 16
C	2	11	100.0	16	6	I27905 Sequence 77
C	3	11	100.0	16	6	I27996 Sequence 16
C	4	11	100.0	17	6	I28051 Sequence 22
C	5	11	100.0	19	6	I28019 Sequence 19
C	6	11	100.0	24	6	A42460 Sequence 9
C	7	11	100.0	25	6	ARI57073 Sequence
C	8	11	100.0	27	6	AX128308 Sequence
C	9	11	100.0	39	6	AX044055 Sequence
C	10	11	100.0	39	6	AX044109 Sequence
C	11	11	100.0	39	6	AX044157 Sequence
C	12	11	100.0	54	6	AR040506 Sequence
C	13	11	100.0	69	6	AR097013 Sequence
C	14	11	100.0	74	6	AR097014 Sequence
C	15	11	100.0	131	6	EL5304 Oryza sativ
C	16	11	100.0	152	9	HUWHDR413
C	17	11	100.0	154	10	S57440S05
C	18	11	100.0	156	11	AU025871
C	19	11	100.0	175	9	S79786
C	20	11	100.0	209	6	AX237100 Sequence
C	21	11	100.0	213	9	AF406781
C	22	11	100.0	220	6	AR034255 Sequence
C	23	11	100.0	225	9	HUWHDR413
C	24	11	100.0	225	9	HUWHDR414
C	25	11	100.0	225	9	HUWHDR4B1
C	26	11	100.0	227	9	HSDBR1125
C	27	11	100.0	228	9	HSU25442
C	28	11	100.0	228	9	HSU25639
C	29	11	100.0	231	9	HSDBR1125
C	30	11	100.0	231	9	HSDBR1125
C	31	11	100.0	233	8	AY022841 Oryza sat
C	32	11	100.0	234	9	HSDBR1125
C	33	11	100.0	234	9	AF028588
C	34	11	100.0	234	9	AF028589
C	35	11	100.0	234	9	AF028590
C	36	11	100.0	234	9	AF144105 Homo sapi
C	37	11	100.0	234	9	AF152844 Homo sapi
C	38	11	100.0	234	9	AF152845 Homo sapi
C	39	11	100.0	234	9	AF278701 Homo sapi
C	40	11	100.0	235	9	HSDBR3MOB
C	41	11	100.0	236	9	AF093411 Homo sapi
C	42	11	100.0	237	9	HSDBR1125
C	43	11	100.0	237	9	HUWHDR414
C	44	11	100.0	237	9	HUWHDR414
C	45	11	100.0	237	11	G19684 human STS A

ALIGNMENTS

1026

RESULT	1	12 bp	DNA	linear	PAT 07-OCT-1996
LOCUS	I24588	Sequence 16 from patent US 5545526.			
DEFINITION	I24588				
ACCESSION	I24588				
VERSION	I24588.1	GI:1604458			
KEYWORDS					
SOURCE	Unknown.				
ORGANISM	Unknown.				
REFERENCE	1 (bases 1 to 12)				
AUTHORS	Baxter-Lowe, L. Ann.				
TITLE	Method for HLA Typing				
JOURNAL	Patent: US 5545526-A 16 13-NOV-1996;				
FEATURES	Location/Qualifiers				
source	1..12	/organism="unknown"			
BASE COUNT	3 a	7 c	0 g	2 t	
ORIGIN					

Query Match 100.0%; Score 11; DB 6; Length 12;  
Best Local Similarity 100.0%; Pred. No. 5.7e+04;

Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 tctcaccacc 11  
|1111111111|  
Db 2 TCTCACCACACC 12

RESULT 2  
LOCUS 127905/c 16 bp DNA  
DEFINITION Sequence 77 from patent US 5567809.  
ACCESSION 127905  
VERSION 127905.1 GI:1818681  
KEYWORDS  
SOURCE Unknown.  
ORGANISM Unknown.  
REFERENCE 1 (bases 1 to 16)  
AUTHORS Apple,R.J., Erlich,H.A., Griffith,R.L. and Scharf,S.J.  
TITLE Methods and reagents for HLA DRbeta DNA typing  
JOURNAL Patent: US 5567809-A 77 22-OCT-1996;  
FEATURES Location/Qualifiers  
source 1..16  
BASE COUNT 2 a 2 c 8 g 4 t  
ORIGIN

Query Match 100.0%; Score 11; DB 6; Length 16;  
Best Local Similarity 100.0%; Pred. No. 5.4e+04;  
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 tctcaccacc 11  
|1111111111|  
Db 13 TCTCACCACACC 3

RESULT 3  
LOCUS 127996/c 16 bp DNA  
DEFINITION Sequence 168 from patent US 5567809.  
ACCESSION 127996  
VERSION 127996.1 GI:1818772  
KEYWORDS  
SOURCE Unknown.  
ORGANISM Unknown.  
REFERENCE 1 (bases 1 to 16)  
AUTHORS Apple,R.J., Erlich,H.A., Griffith,R.L. and Scharf,S.J.  
TITLE Methods and reagents for HLA DRbeta DNA typing  
JOURNAL Patent: US 5567809-A 168 22-OCT-1996;  
FEATURES Location/Qualifiers  
source 1..16  
BASE COUNT 2 a 1 c 9 g 4 t  
ORIGIN

Query Match 100.0%; Score 11; DB 6; Length 16;  
Best Local Similarity 100.0%; Pred. No. 5.4e+04;  
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 tctcaccacc 11  
|1111111111|  
Db 13 TCTCACCACACC 3

RESULT 4  
LOCUS 128051/c 17 bp DNA  
DEFINITION Sequence 223 from patent US 5567809.  
ACCESSION 128051  
VERSION 128051.1 GI:1818827

KEYWORDS  
SOURCE Unknown.  
ORGANISM Unknown.  
REFERENCE 1 (bases 1 to 17)  
AUTHORS Apple,R.J., Erlich,H.A., Griffith,R.L. and Scharf,S.J.  
TITLE Methods and reagents for HLA DRbeta DNA typing  
JOURNAL Patent: US 5567809-A 223 22-OCT-1996;  
FEATURES Location/Qualifiers  
source 1..17  
BASE COUNT 2 a 2 c 9 g 4 t  
ORIGIN

Query Match 100.0%; Score 11; DB 6; Length 17;  
Best Local Similarity 100.0%; Pred. No. 5.4e+04;  
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 tctcaccacc 11  
|1111111111|  
Db 14 TCTCACCACACC 4

RESULT 5  
LOCUS 128019/c 19 bp DNA  
DEFINITION Sequence 191 from patent US 5567809.  
ACCESSION 128019  
VERSION 128019.1 GI:1818795  
KEYWORDS  
SOURCE Unknown.  
ORGANISM Unknown.  
REFERENCE 1 (bases 1 to 19)  
AUTHORS Apple,R.J., Erlich,H.A., Griffith,R.L. and Scharf,S.J.  
TITLE Methods and reagents for HLA DRbeta DNA typing  
JOURNAL Patent: US 5567809-A 191 22-OCT-1996;  
FEATURES Location/Qualifiers  
source 1..19  
BASE COUNT 3 a 2 c 9 g 5 t  
ORIGIN

Query Match 100.0%; Score 11; DB 6; Length 19;  
Best Local Similarity 100.0%; Pred. No. 5.3e+04;  
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 tctcaccacc 11  
|1111111111|  
Db 15 TCTCACCACACC 5

RESULT 6  
LOCUS A42460 24 bp DNA  
DEFINITION Sequence 9 from Patent WO9503331.  
ACCESSION A42460  
VERSION A42460.1 GI:2297917  
KEYWORDS  
SOURCE unidentified.  
ORGANISM unidentified.  
REFERENCE 1 (bases 1 to 24)  
AUTHORS Whitaker,J.L. and Morten,J.E.  
TITLE HUMAN MHC CLASS II DOUBLE TRANSGENE AND USES  
JOURNAL Patent: WO 9503331-A 9 02-FEB-1995;  
COMMENT ZENECA LTD (GB)  
Other publication AU 7231494 950220  
FEATURES Other publication GB 2280186 950125.  
source 1..24

BASE COUNT 6 a 10 c 4 g 4 t  
ORIGIN

Query Match 100.0%; Score 11; DB 6; Length 24;  
Best Local Similarity 100.0%; Pred. No. 5e+04;  
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 tctcccaacc 11  
|||||  
Db 11 TCTCACCAACC 21

RESULT 7  
ARI57073/c ARI57073 25 bp DNA linear PAT 08-AUG-2001  
LOCUS  
DEFINITION Sequence 10 from patent US 6242588.  
ACCESSION ARI57073  
VERSION ARI57073.1 GI:15125777  
KEYWORDS  
SOURCE unknown.  
ORGANISM unknown.  
REFERENCE 1 (bases 1 to 25)  
AUTHORS Sheppard,P.O., Piddington,C.S. and Ellsworth,J.L.  
TITLE Testis specific glycoprotein zpep10  
JOURNAL Patent: US 6242588-A 10 05-JUN-2001;  
FEATURES Location/Qualifiers  
source 1..25

BASE COUNT 4 a 5 c 7 g 9 t  
ORIGIN

Query Match 100.0%; Score 11; DB 6; Length 25;  
Best Local Similarity 100.0%; Pred. No. 5e+04;  
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 tctcccaacc 11  
|||||  
Db 22 TCTCACCAACC 12

RESULT 8  
AX128308/c AX128308 27 bp DNA linear PAT 15-MAY-2001  
LOCUS  
DEFINITION Sequence 75 from Patent W00130992.  
ACCESSION AX128308  
VERSION AX128308.1 GI:14134829  
KEYWORDS  
SOURCE synthetic construct.  
ORGANISM synthetic construct.  
REFERENCE 1 (bases 1 to 27)  
AUTHORS Koike,C.  
TITLE gta11-3 galactosyltransferase gene and promoter  
JOURNAL Patent: W0 0130992-A 75 03-MAY-2001;  
UNIV. PITTSBURGH OF THE COMMONWEALTH SYSTEM OF HIGHER EDUCATION  
(US)  
FEATURES Location/Qualifiers  
source 1..27  
/organism="synthetic construct"  
/db\_xref="taxon:32630"  
/note="Primer for identifying murine exons 6 and 7"

BASE COUNT 7 a 4 c 10 g 6 t  
ORIGIN

Query Match 100.0%; Score 11; DB 6; Length 27;  
Best Local Similarity 100.0%; Pred. No. 4.9e+04;  
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 tctcccaacc 11  
|||||  
Db 18 TCTCACCAACC 8

RESULT 9  
AX044055/c AX044055 39 bp DNA linear PAT 24-NOV-2000  
LOCUS  
DEFINITION Sequence 9 from Patent W00066748.  
ACCESSION AX044055  
VERSION AX044055.1 GI:11342933  
KEYWORDS  
SOURCE synthetic construct.  
ORGANISM synthetic construct.  
REFERENCE 1 (bases 1 to 39)  
AUTHORS Hawkes,T.R., Warner,S.A., Andrews,C.J., Bachoo,S. and  
TITLE Herbicide resistant plants  
JOURNAL Patent: W0 0066748-A 9 09-NOV-2000;  
ZENDECA LIMITED (GB)  
FEATURES Location/Qualifiers  
source 1..39  
/organism="synthetic construct"  
/db\_xref="taxon:32630"  
/note="Primer"

BASE COUNT 7 a 9 c 15 g 8 t  
ORIGIN

Query Match 100.0%; Score 11; DB 6; Length 39;  
Best Local Similarity 100.0%; Pred. No. 4.6e+04;  
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 tctcccaacc 11  
|||||  
Db 29 TCTCACCAACC 19

RESULT 10  
AX044109/c AX044109 39 bp DNA linear PAT 24-NOV-2000  
LOCUS  
DEFINITION Sequence 9 from Patent W00066747.  
ACCESSION AX044109  
VERSION AX044109.1 GI:11342987  
KEYWORDS  
SOURCE synthetic construct.  
ORGANISM synthetic construct.  
REFERENCE 1 (bases 1 to 39)  
AUTHORS Hawkes,T.R., Warner,S.A., Andrews,C.J., Bachoo,S. and  
TITLE Herbicide resistant plants  
JOURNAL Patent: W0 0066747-A 9 09-NOV-2000;  
ZENDECA LIMITED (GB)  
FEATURES Location/Qualifiers  
source 1..39  
/organism="synthetic construct"  
/db\_xref="taxon:32630"  
/note="Primer"

BASE COUNT 7 a 9 c 15 g 8 t  
ORIGIN

Query Match 100.0%; Score 11; DB 6; Length 39;  
Best Local Similarity 100.0%; Pred. No. 4.6e+04;  
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 tctcccaacc 11  
|||||  
Db 29 TCTCACCAACC 19

RESULT 11  
AX044157/c 39 bp DNA linear PAT 24-NOV-2000  
LOCUS AX044157  
DEFINITION Sequence 9 from Patent WO0066746.  
ACCESSION AX044157  
VERSION AX044157.1 GI:11343035  
KEYWORDS  
SOURCE  
ORGANISM  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
FEATURES  
source  
1. .39  
/organism="synthetic construct"  
/db\_xref="taxon:32630"  
/note="Primer"  
BASE COUNT 7 a 9 c 15 g 8 t  
ORIGIN

Query Match 100.0%; Score 11; DB 6; Length 39;  
Best Local Similarity 100.0%; Pred. No. 4.6e+04;  
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 tctcaccac 11  
DB 29 TCTCACCAC 19  
|||||

RESULT 12  
AR040506/c 54 bp DNA linear PAT 29-SEP-1999  
LOCUS AR040506  
DEFINITION Sequence 1354 from patent US 5807743.  
ACCESSION AR040506  
VERSION AR040506.1 GI:5959869  
KEYWORDS  
SOURCE  
ORGANISM  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
FEATURES  
source  
1. .54  
/organism="unknown"  
BASE COUNT 15 a 9 c 18 g 12 t  
ORIGIN

Query Match 100.0%; Score 11; DB 6; Length 54;  
Best Local Similarity 100.0%; Pred. No. 4.3e+04;  
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 tctcaccac 11  
DB 11 TCTCACCAC 1  
|||||

RESULT 13  
AR097013/c 69 bp DNA linear PAT 14-FEB-2001  
LOCUS AR097013  
DEFINITION Sequence 76 from patent US 6071515.  
ACCESSION AR097013  
VERSION AR097013.1 GI:12805743  
KEYWORDS  
SOURCE  
ORGANISM  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
FEATURES  
source  
Unknown.

ORGANISM Unknown.  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
FEATURES  
source  
1. .69  
/organism="unknown"  
BASE COUNT 16 a 16 c 24 g 13 t  
ORIGIN

Query Match 100.0%; Score 11; DB 6; Length 69;  
Best Local Similarity 100.0%; Pred. No. 4.1e+04;  
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 tctcaccac 11  
DB 41 TCTCACCAC 31  
|||||

RESULT 14  
AR097014 74 bp DNA linear PAT 14-FEB-2001  
LOCUS AR097014  
DEFINITION Sequence 77 from patent US 6071515.  
ACCESSION AR097014  
VERSION AR097014.1 GI:12805744  
KEYWORDS  
SOURCE  
ORGANISM  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
FEATURES  
source  
1. .74  
/organism="unknown"  
BASE COUNT 14 a 25 c 17 g 18 t  
ORIGIN

Query Match 100.0%; Score 11; DB 6; Length 74;  
Best Local Similarity 100.0%; Pred. No. 4.1e+04;  
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 tctcaccac 11  
DB 32 TCTCACCAC 42  
|||||

RESULT 15  
E15304/c 131 bp DNA linear PAT 28-JUL-1999  
LOCUS E15304  
DEFINITION Oryza sativa microsatellite marker.  
ACCESSION E15304  
VERSION E15304.1 GI:5709987  
KEYWORDS  
SOURCE  
ORGANISM  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
COMMENT  
OS  
PN  
PD

1 (bases 1 to 131)  
Akagi, H., Fujimura, T., Yokozeki, S. and Inagaki, A.  
MICROSATELLITE MARKER FOR DISCRIMINATING RICE PLANT CULTIVAR AND  
TEST FOR PURITY OF PLANT SEED  
Patent: JP 1998057073-A 28 03-MAR-1998;  
MITSUI PETROCHEM IND LTD  
OS Oryza sativa  
PN JP 1998057073-A/28  
PD 03-MAR-1998

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;  
Ehrhartoideae; Oryzaceae; Oryza.

PF 25-FEB-1997 JP 1997040226  
PR 13-JUN-1996 JP 96P 152657  
PI AKAGI HIROMORI, FUJIMURA TATSURO, YOKOZAKI SUKEYOSHI, PI  
INAGAKI AKIKO  
PC C12N15/09,C12Q1/68;  
CC strandedness: Double;  
CC topology: Linear;  
FH key Location/Qualifiers  
FT source 1..131  
FT repeat\_region 1..131  
FT Location/Qualifiers  
1..131  
/organism="Oryza sativa"  
/db\_xref="taxon:4530"  
BASE COUNT 26 a 27 c 63 g 15 t  
ORIGIN

Query Match 100.0%; Score 11; DB 6; Length 131;  
Best Local Similarity 100.0%; Pred. No. 3.7e+04;  
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 tctcaccacc 11  
|||||  
Db 41 TCTCACCAC 31

Search completed: July 29, 2002, 23:55:08  
Job time: 7024 sec



GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 30, 2002, 00:01:19 ; Search time 285.14 Seconds  
(without alignments)  
66.234 Million cell updates/sec

Title: US-09-530-663B-17  
Perfect score: 11  
Sequence: 1 tctcaccacac 11

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 1736436 segs, 858457221 residues

Total number of hits satisfying chosen parameters: 3472872

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

N.GeneSeq\_032802:\*  
1: /SIDSL/gcgdata/geneSeq/geneSeqn-emb1/NA1980.DAT:\*  
2: /SIDSL/gcgdata/geneSeq/geneSeqn-emb1/NA1981.DAT:\*  
3: /SIDSL/gcgdata/geneSeq/geneSeqn-emb1/NA1982.DAT:\*  
4: /SIDSL/gcgdata/geneSeq/geneSeqn-emb1/NA1983.DAT:\*  
5: /SIDSL/gcgdata/geneSeq/geneSeqn-emb1/NA1984.DAT:\*  
6: /SIDSL/gcgdata/geneSeq/geneSeqn-emb1/NA1985.DAT:\*  
7: /SIDSL/gcgdata/geneSeq/geneSeqn-emb1/NA1986.DAT:\*  
8: /SIDSL/gcgdata/geneSeq/geneSeqn-emb1/NA1987.DAT:\*  
9: /SIDSL/gcgdata/geneSeq/geneSeqn-emb1/NA1988.DAT:\*  
10: /SIDSL/gcgdata/geneSeq/geneSeqn-emb1/NA1989.DAT:\*  
11: /SIDSL/gcgdata/geneSeq/geneSeqn-emb1/NA1990.DAT:\*  
12: /SIDSL/gcgdata/geneSeq/geneSeqn-emb1/NA1991.DAT:\*  
13: /SIDSL/gcgdata/geneSeq/geneSeqn-emb1/NA1992.DAT:\*  
14: /SIDSL/gcgdata/geneSeq/geneSeqn-emb1/NA1993.DAT:\*  
15: /SIDSL/gcgdata/geneSeq/geneSeqn-emb1/NA1994.DAT:\*  
16: /SIDSL/gcgdata/geneSeq/geneSeqn-emb1/NA1995.DAT:\*  
17: /SIDSL/gcgdata/geneSeq/geneSeqn-emb1/NA1996.DAT:\*  
18: /SIDSL/gcgdata/geneSeq/geneSeqn-emb1/NA1997.DAT:\*  
19: /SIDSL/gcgdata/geneSeq/geneSeqn-emb1/NA1998.DAT:\*  
20: /SIDSL/gcgdata/geneSeq/geneSeqn-emb1/NA1999.DAT:\*  
21: /SIDSL/gcgdata/geneSeq/geneSeqn-emb1/NA2000.DAT:\*  
22: /SIDSL/gcgdata/geneSeq/geneSeqn-emb1/NA2001A.DAT:\*  
23: /SIDSL/gcgdata/geneSeq/geneSeqn-emb1/NA2001B.DAT:\*  
24: /SIDSL/gcgdata/geneSeq/geneSeqn-emb1/NA2002.DAT:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	11	100.0	12	17	AAx79382
2	11	100.0	12	17	AAx79382
3	11	100.0	12	17	AAx79382
4	11	100.0	12	22	AAx79382
5	11	100.0	12	22	AAx79382
6	11	100.0	16	13	AAx79382
7	11	100.0	16	13	AAx79382
8	11	100.0	17	13	AAx79382
9	11	100.0	17	24	AAx79382

C	10	11	100.0	18	22	AAx79486
C	11	11	100.0	18	24	AAx79486
C	12	11	100.0	18	24	AAx79486
C	13	11	100.0	19	11	AAx79486
C	14	11	100.0	19	11	AAx79486
C	15	11	100.0	19	17	AAx79486
C	16	11	100.0	19	19	AAx79486
C	17	11	100.0	19	22	AAx79486
C	18	11	100.0	22	21	AAx79486
C	19	11	100.0	24	16	AAx79486
C	20	11	100.0	25	21	AAx79486
C	21	11	100.0	25	22	AAx79486
C	22	11	100.0	27	22	AAx79486
C	23	11	100.0	39	21	AAx79486
C	24	11	100.0	39	21	AAx79486
C	25	11	100.0	39	21	AAx79486
C	26	11	100.0	52	19	AAx79486
C	27	11	100.0	122	21	AAx79486
C	28	11	100.0	131	19	AAx79486
C	29	11	100.0	135	22	AAx79486
C	30	11	100.0	170	22	AAx79486
C	31	11	100.0	203	22	AAx79486
C	32	11	100.0	203	22	AAx79486
C	33	11	100.0	203	22	AAx79486
C	34	11	100.0	203	22	AAx79486
C	35	11	100.0	209	22	AAx79486
C	36	11	100.0	220	16	AAx79486
C	37	11	100.0	220	21	AAx79486
C	38	11	100.0	240	17	AAx79486
C	39	11	100.0	240	21	AAx79486
C	40	11	100.0	264	21	AAx79486
C	41	11	100.0	268	13	AAx79486
C	42	11	100.0	268	20	AAx79486
C	43	11	100.0	269	13	AAx79486
C	44	11	100.0	269	13	AAx79486
C	45	11	100.0	269	17	AAx79486

#### ALIGNMENTS

RESULT	1
ID	AAx79382/c
AAx79382	standard; DNA; 12 BP.
AC	AAx79382:
XX	
XX	17-AUG-1999 (first entry)
DT	
XX	
DE	HLA-DR typing probe G86.
XX	
KW	Tissue typing: human leukocyte antigen: HLA: MHC: donor: allele; PCR:
KW	major histocompatibility complex; bone marrow transplant; primer:
KW	amplification; polymerase chain reaction; probe; polymorphism;
KW	sequence-specific oligonucleotide probe hybridisation; ss.
XX	
OS	Synthetic.
XX	
PN	US5468611-A.
XX	
PD	21-NOV-1995.
XX	
PF	08-APR-1993; 93US-0045530.
XX	
PR	27-JUN-1990; 90US-0544218.
XX	
PA	08-APR-1993; 93US-0045530.
XX	
PI	(BLOO-) BLOOD CENT RES FOUND INC.
XX	
XX	Baxter-Lowe LA, Gorski JA;
DR	WPI; 1996-010091/01.
XX	

Glucanase genomic  
Human HLA genotype  
Human HLA genotype  
Oligonucleotide pr  
HLA-DR beta sub-ty  
HLA-DR typing prob  
Probe VAL86 used t  
HLA-DR typing prob  
Gene typing PCR pr  
Oligonucleotide pr  
Human testis speci  
Human testis-speci  
Mouse alpha-1,3 ga  
Rice BPSPS PCR pri  
Primer G1. Synthe  
Canine IL-2 recept  
Human secreted pro  
Microsatellite mar  
HLA-DR beta allele  
Human secreted pro  
Human foetal liver  
Human brain expres  
Human bone marrow  
Probe #19586 used  
Human haematologic  
Spleen necrosis vi  
Eucaryotic gene ex  
HLA-DR beta allele  
HLA-DR beta allele  
Arabidopsis thaila  
DRI. Synthetic.  
HLA allele DRB1\*04  
DRB1\*1105 allele.  
DRB1\*1305 allele.  
HLA-DR beta allele

PT Improved method for HLA typing - by DNA amplification and  
PT sequence-specific oligo:nucleotide hybridisation, used to select  
PS bone marrow donors  
XX Disclosure: Column 19-20; 20pp; English.  
XX  
CC A novel method of typing the human leukocyte antigen (HLA) of the major  
CC histocompatibility complex (MHC), esp. for typing donors for bone marrow  
CC transplants, involves determining if the donor tissue HLA-DR alleles are  
CC selected from the gp.: HLA-DRB1\*08, DR12a,b, DR3a,n, DR5a-e, DRB1\*01,  
CC DR6a, DRB1\*03, DRB1\*04, DR1, DR9, DR2a-c, DR2a-d, DR10 and  
CC DR11a-c. The method uses PCR to amplify these regions followed by  
CC sequence-specific oligonucleotide probe hybridisation (SSOPH) using the  
CC probes AAV19365-179429. SSOPH allows detection of polymorphisms that  
CC predict differences at a single amino acid level thus reducing errors  
CC and improving the chance of successfully matching tissues.  
XX  
SQ Sequence 12 BP: 2 A; 0 C; 7 G; 3 T; 0 other;

Query Match 100.0%; Score 11; DB 17; Length 12;  
Best Local Similarity 100.0%; Pred. No. 1.9e+03;  
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 tctcaccacc 11  
|||||  
DB 11 TCTCACCACC 1

RESULT 2  
AAT41819  
ID AAT41819 standard; DNA: 12 BP.  
XX  
AC AAT41819;  
XX  
DT 18-DEC-1996 (first entry)  
XX  
DE HLA allele, HLA-DRB1\*08, \*12 and \*1404 resolution probe, G86.  
XX  
XX Human leukocyte antigen; HLA: allele; HLA-DR\*08; HLA-DR\*12; locus B1;  
KM polymorphism; amplify; conserved region; detection; primer; probe;  
KM tissue matching; identifying disease susceptibility; ss.  
XX  
OS Synthetic.  
XX  
PN US5545526-A.  
XX  
PD 13-AUG-1996.  
XX  
PF 27-JUN-1990; 90US-0544218.  
XX  
PR 01-MAR-1993; 93US-0025038.  
PR 27-JUN-1990; 90US-0544218.  
XX  
PA (BLOO-) BLOOD CENT RES FOUND INC.  
XX  
PI Baxter-Lowe LA;  
XX  
XX WPI; 1996-383664/38.  
XX  
PT Human leukocyte antigen typing of tissue samples - using  
PT allele-specific amplification to distinguish allele pairs  
XX  
XX Example 1; Column 19; 24pp; English.  
XX  
CC The sequences given in AAT41811-20 represent probes which were used to  
CC resolve the human leukocyte antigen (HLA) DRB1 alleles, DRB1\*08, \*12  
CC and \*1404. This probe sequence hybridises to the gly86 coding region  
CC found in alleles \*0801, \*0803 and 0805. These probes may be used  
CC in the method of invention which concerns HLA typing of a sample for an  
CC unknown pair of alleles. The pair of alleles comprises one of two known  
CC types which have the same overall set of polymorphisms but have a  
CC different distribution of polymorphisms between their two alleles. The

CC method comprises selectively amplifying the DNA of just one allele of  
CC the unknown pair and analysing the amplified DNA to determine which  
CC polymorphisms are present in that allele, and therefore assigning the  
CC unknown pair to the known type having that allele. The method comprises  
CC three test stages. The first stage is to establish the number of  
CC alleles present in each sample. Primers corresponding to fairly well  
CC conserved regions of a locus will increase the likelihood that unknown  
CC alleles will be amplified and potentially detected by hybridisation with  
CC a probe. In the second stage, the group or basic type identified  
CC determines which set of allele specific primers will be used. The first  
CC of the two primers comprises an opt. labeled sequence common to each  
CC allele of the group identified in the first stage but different from  
CC other groups identified in stage one. The second primer may be a  
CC mixture of different labeled primers, complementary to two or more  
CC sequences within the group, or the amplification may be performed with  
CC only one second primer to detect the presence of a single group of  
CC alleles. In the third stage the specific allele is determined. This  
CC may be done by amplification or hybridisation using a radiolabelled  
CC probe. The method may be used for tissue matching, identifying disease  
CC susceptibility, etc. The method of the invention esp. distinguishes  
CC between DOB1\*0304/DOB1\*03032 and DOB1\*0301/DOB1\*0302.  
XX  
SQ Sequence 12 BP: 3 A; 7 C; 0 G; 2 T; 0 other;

Query Match 100.0%; Score 11; DB 17; Length 12;  
Best Local Similarity 100.0%; Pred. No. 1.9e+03;  
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 tctcaccacc 11  
|||||  
DB 2 tctcaccacc 12

RESULT 3  
AAV16578/C  
ID AAV16578 standard; DNA: 12 BP.  
XX  
AC AAV16578;  
XX  
DT 12-JUN-1998 (first entry)  
XX  
DE Probe G86 used to identify HLA-DR sequences.  
XX  
XX DR region; major histocompatibility complex; HLA-DR; HLA-typing;  
KM HLA-DR beta consensus sequence; allelic polymorphism;  
KM HLA-DR beta-allelic polymorphism; probe; bone marrow; transplant; ss.  
XX  
XX Synthetic.  
XX  
OS Homo sapiens.  
XX  
PN US5702885-A.  
XX  
PD 30-DEC-1997.  
XX  
PF 08-APR-1993; 93US-0057957.  
XX  
PR 27-JUN-1990; 90US-0544218.  
XX  
PA (BLOO-) BLOOD CENT RES FOUND INC.  
XX  
PI Baxter-Lowe LA, Gorski JA;  
XX  
XX WPI; 1998-076408/07.  
XX  
PT Oligo:nucleotide probes and primers and methods for HLA typing -  
PT particularly for tissue typing for bone marrow transplants  
XX  
XX Disclosure: Column 19; 20pp; English.  
XX  
CC Probes AAV16561-624 are used to identify differences in the DR region of  
CC human major histocompatibility complex (HLA-DR). The specification  
CC describes a method for HLA-typing, which includes an oligonucleotide



CC probe which undergoes sequence-specific hybridisation with an HLA-DR  
 CC beta consensus sequence at positions 61-64. The probe contains a  
 CC labelling substance other than a nucleotide sequence, which facilitates  
 CC detection of the probe. The HLA sequence of a subject is PCR amplified,  
 CC and a probe that recognises an allelic polymorphism at a selected HLA  
 CC locus is contacted with the amplified product. This first probe  
 CC recognises a HLA-DR beta-allelic polymorphism. A second (different)  
 CC probe is brought into contact with a second sample of the amplified DNA  
 CC in a separate reaction, and hybridisation detected. The probes and  
 CC primers are used for HLA typing, e.g. for tissue, especially bone  
 CC marrow, transplants.  
 XX  
 SQ Sequence 12 BP; 2 A; 0 C; 7 G; 3 T; 0 other;

Query Match 100.0%; Score 11; DB 19; Length 12;  
 Best Local Similarity 100.0%; Pred. No. 1.9e+03;  
 Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 tctcaccacc 11  
 |||||  
 DB 11 TCTCACCACC 1

RESULT 4  
 AAF92638/C  
 ID AAF92638 standard; DNA: 12 BP.

XX AAF92638;  
 AC  
 XX 16-MAY-2001 (first entry)  
 DT  
 XX  
 DE HLA-DR typing probe #18.

XX Human; leukocyte antigen; HLA; typing; sequence specific probe;  
 KW SSOPH; ss.  
 XX  
 OS Homo sapiens.

XX US6194147-B1.  
 PN

XX 27-FEB-2001.  
 PD

XX 30-DEC-1997; 97US-0000805.  
 PF

XX 27-JUN-1990; 90US-0544218.  
 PR 08-APR-1993; 93US-0057957.

XX (BLOO-) BLOOD CENT RES FOUND INC.  
 PA

PI Baxter-Lowe LA, Gorski JA;  
 DR

XX WPI: 2001-217923/22.  
 DR

PT Human leukocyte antigen typing by amplifying a sample followed by  
 PT sequence specific oligonucleotide hybridization with labeled  
 PT oligonucleotide probes that hybridize with a series of known control  
 PT DNA sequences -

XX Disclosure; Column 11-14; 16pp; English.  
 PS

CC The present invention relates to human leukocyte antigen (HLA) typing.  
 CC The method involves detecting polymorphic residues by sequence  
 CC specific oligonucleotide probe hybridization (SSOPH) with labeled  
 CC oligonucleotide probes.  
 XX

SQ Sequence 12 BP; 2 A; 0 C; 7 G; 3 T; 0 other;

Query Match 100.0%; Score 11; DB 22; Length 12;  
 Best Local Similarity 100.0%; Pred. No. 1.9e+03;  
 Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 tctcaccacc 11  
 |||||  
 DB 11 TCTCACCACC 1

RESULT 5  
 AAQ26124/C  
 ID AAQ26124 standard; DNA: 16 BP.

XX AAQ26124;  
 AC

DT 04-JAN-1993 (first entry)  
 DT

DE HLA-DR beta sub-type tailed probe DRB15 hybridising region.  
 DE

XX Tissue typing; identity determination; disease susceptible; ss.  
 KW

XX Synthetic.  
 OS

XX WO9210589-A.  
 PN

XX 25-JUN-1992.  
 PD

XX 06-DEC-1991; 91WO-US09294.  
 PF

XX 06-DEC-1990; 90US-0623098.  
 PR

XX (HOFF) HOFFMANN LA ROCHE & CO AG F.  
 PA

XX Apple RJ, Begovich AB, Bugawan T, Erlich HA, Griffith RL;  
 PI Scharf SJ;  
 PI

DR WPI: 1992-234644/28.  
 DR

XX Method for determining HLA-DR beta sub-type in DNA sample -  
 PT comprises amplification and hybridisation with probes and  
 PT primers, useful in tissue typing  
 PT

XX Example; Page 37; 90pp; English.  
 PS

XX The sequence is that of the hybridising region of tailed probe DRB15 for  
 CC use in a method for determining HLA-DR beta sub-type in a nucleic acid  
 CC sample. The method allows specific nucleic acid sequences of the second  
 CC exon of HLA-DR beta genes to be amplified then probed for identification  
 CC of polymorphic sequences. The amplified DNA is useful for typing  
 CC homozygous or heterozygous samples from a variety of sources and for  
 CC detecting allelic variants not distinguishable by serological methods.  
 CC The typing system can be used in a reverse dot blot format which is  
 CC simple and rapid to perform, produces detectable signals in minutes and  
 CC can be utilised in tissue typing, determination of individual identity  
 CC and identifying disease susceptible individuals. Preliminary testing  
 CC shows that the probe is more preferred than others. The probe is  
 CC used with the HRP-labelled, untailled probe CRX36.  
 CC See also AAQ26092-Q26367.  
 CC

SQ Sequence 16 BP; 2 A; 2 C; 8 G; 4 T; 0 other;

Query Match 100.0%; Score 11; DB 13; Length 16;  
 Best Local Similarity 100.0%; Pred. No. 1.9e+03;  
 Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 tctcaccacc 11  
 |||||  
 DB 13 TCTCACCACC 3

RESULT 6  
 AAQ26196/C  
 ID AAQ26196 standard; DNA: 16 BP.

XX AAQ26196;  
 AC  
 XX

```

DT 04-JAN-1993 (first entry)
XX
XX HLA-DR beta sub-type tailed probe DRB92 hybridising region.
DE
XX Tissue typing; identity determination; disease susceptible; ss.
XX
XX Synthetic.
OS
XX WO9210589-A.
PN
XX
XX 25-JUN-1992.
PD
XX
XX 06-DEC-1991; 91WO-US09294.
PF
XX
XX 06-DEC-1990; 90US-0623098.
PR
XX
XX (HOFF ) HOFFMANN LA ROCHE & CO AG F.
PA
XX Apple RJ, Begovich AB, Bugawan T, Erlich HA, Griffith RL,
PI Scharf SJ;
DR WPI; 1992-234644/28.
XX
XX Method for determining HLA-DR beta sub-type in DNA sample -
PT comprises amplification and hybridisation with probes and
PT primers, useful in tissue typing
PS
XX Example; Page 39; 90pp; English.
CC The sequence is that of the hybridising region of tailed probe DRB92 for
CC use in a method for determining HLA-DR beta sub-type in a nucleic acid
CC sample. The method allows specific nucleic acid sequences of the second
CC exon of HLA-DR beta genes to be amplified then probed for identification
CC of polymorphic sequences. The amplified DNA is useful for identifying
CC homozygous or heterozygous samples from a variety of sources and for
CC detecting allelic variants not distinguishable by serological methods.
CC The typing system can be used in a reverse dot blot format which is
CC simple and rapid to perform, produces detectable signals in minutes and
CC can be utilised in tissue typing, determination of individual identity
CC and identifying disease susceptible individuals.
CC See also AAQ26092-Q26367.
CC
XX
XX Sequence 16 BP; 2 A; 1 C; 9 G; 4 T; 0 other;
SQ

Query Match 100.0%; Score 11; DB 13; Length 16;
Best Local Similarity 100.0%; Pred. No. 1.9e+03;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 tctcaccacc 11
   |||||
DB 13 TCTCACCACAC 3

RESULT 7
ABL31303/c
ID ABL31303 standard; DNA; 16 BP.
XX
XX ABL31303;
AC
XX
XX 21-MAR-2002 (first entry)
DT
XX
XX Human HLA genotyping oligonucleotide SEQ ID NO 792.
DE
XX
XX Human; human leukocyte antigen; HLA; genotype; polymorphism;
KW immunogenetic; transplantation; genetic disease; ss.
XX
XX Homo sapiens.
OS
XX WO200192572-A1.
PN
XX
XX 06-DEC-2001.
PD
XX

```

```

PF 01-JUN-2001; 2001WO-JP04662.
XX
XX 01-JUN-2000; 2000JP-0164798.
PR
XX
XX (NISN ) NISSHINO IND INC.
PA (SYST-) SYSTEM RES INC.
XX
XX Inoko H, Kagiya T, Ichihara T, Matsumura Y, Moriya S, Nishida M;
PI WPI; 2002-122074/16.
XX
XX Human leukocyte antigen (HLA) typing, useful for judging HLA genotypes
PT of individuals e.g. by determining immunogenetic differences when
PT transplanting between them -
PS
XX Claim 10; Page 243; 345pp; Japanese.
CC The invention relates to a typing kit for judging human leukocyte antigen
CC (HLA) genotype of a sample by hybridising a substrate on which 10-24 base
CC oligonucleotides (ABL30512-ABL31809) originating in the sequences of
CC genes e.g. belonging to HLA class I antigens on human genome and
CC containing gene polymorphisms as alloantigens have been immobilised as
CC primers for amplification of cleaved nucleic acids relating to gene
CC polymorphisms. The method is useful for judging HLA genotypes of
CC individuals by determining immunogenetic differences before transplanting
CC between them, providing genetic information to decide compatibility of
CC organ and tissue for transplantation e.g. of bone marrow, kidney, liver,
CC pancreas, Langerhans islet in pancreas and cornea, susceptibility
CC diagnosis of genetic diseases and identifying individuals.
CC
XX
XX Sequence 16 BP; 2 A; 1 C; 9 G; 4 T; 0 other;
SQ

Query Match 100.0%; Score 11; DB 24; Length 16;
Best Local Similarity 100.0%; Pred. No. 1.9e+03;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 tctcaccacc 11
   |||||
DB 13 TCTCACCACAC 3

RESULT 8
AAQ26251/c
ID AAQ26251 standard; DNA; 17 BP.
XX
XX AAQ26251;
AC
XX
XX 04-JAN-1993 (first entry)
DT
XX
XX HLA-DR beta sub-type tailed probe DRB147 hybridising region.
DE
XX
XX Tissue typing; identity determination; disease susceptible; ss.
KW
XX
XX Synthetic.
OS
XX WO9210589-A.
PN
XX
XX 25-JUN-1992.
PD
XX
XX 06-DEC-1991; 91WO-US09294.
PF
XX
XX 06-DEC-1990; 90US-0623098.
PR
XX
XX (HOFF ) HOFFMANN LA ROCHE & CO AG F.
PA
XX Apple RJ, Begovich AB, Bugawan T, Erlich HA, Griffith RL;
PI Scharf SJ;
DR WPI; 1992-234644/28.
XX
XX Method for determining HLA-DR beta sub-type in DNA sample -
PT comprises amplification and hybridisation with probes and

```

PT primers, useful in tissue typing  
XX  
PS Example; Page 41; 90pp; English.  
XX  
CC The sequence is that of the hybridizing region of tailed probe DRB147  
CC for use in a method for determining HLA-DR beta sub-type in a nucleic  
CC acid sample. The method allows specific nucleic acid sequences of the  
CC second exon of HLA-DR beta genes to be amplified then probed for  
CC identification of polymorphic sequences. The amplified DNA is useful for  
CC typing homozygous or heterozygous samples from a variety of sources and  
CC for detecting allelic variants not distinguishable by serological  
CC methods. The typing system can be used in a reverse dot blot format which  
CC is simple and rapid to perform, produces detectable signals in minutes  
CC and can be utilised in tissue typing, determination of individual  
CC identity and identifying disease susceptible individuals. It has not yet  
CC been tested. See also AAQ26092-Q26367.  
CC  
SQ Sequence 17 BP; 2 A; 2 C; 9 G; 4 T; 0 other;  
SQ  
Query Match 100.0%; Score 11; DB 13; Length 17;  
Best Local Similarity 100.0%; Pred. No. 1.9e+03;  
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 tctcaccacc 11  
|||||  
DB 14 TCTCACCACACC 4  
RESULT 9  
ABL30817/C  
ID ABL30817 standard; DNA; 17 BP.  
XX  
AC ABL30817;  
XX  
DT 21-MAR-2002 (first entry)  
XX  
DE Human HLA genotyping oligonucleotide SEQ ID NO 306.  
XX  
KW Human leukocyte antigen; HLA; genotype; polymorphism;  
KW immunogenetic; transplantation; genetic disease; ss.  
XX  
OS Homo sapiens.  
XX  
PN WO200192572-A1.  
XX  
PD 06-DEC-2001.  
XX  
PF 01-JUN-2001; 2001WO-JP04662.  
XX  
PR 01-JUN-2000; 2000JP-0164798.  
XX  
PA (NISN ) NISSHINBO IND INC.  
PA (SYST-) SYSTEM RES INC.  
PI Inoko H, Kagiya T, Ichihara T, Matsumura Y, Moriya S, Nishida M;  
XX  
DR WPI; 2002-122074/16.  
XX  
PT Human leukocyte antigen (HLA) typing, useful for judging HLA genotypes  
PT of individuals e.g. by determining immunogenetic differences when  
PT transplanting between them -  
XX  
PS Claim 10; Page 151; 345pp; Japanese.  
XX  
CC The invention relates to a typing kit for judging human leukocyte antigen  
CC (HLA) genotype of a sample by hybridising a substrate on which 10-24 base  
CC oligonucleotides (ABL30512-ABL31809) originating in the sequences of  
CC genes e.g. belonging to HLA class I antigens on human genome and  
CC containing gene polymorphisms as alloantigens have been immobilised as  
CC primers for amplification of cleaved nucleic acids relating to gene  
CC polymorphisms. The method is useful for judging HLA genotypes of  
CC individuals by determining immunogenetic differences before transplanting

CC between them, providing genetic information to decide compatibility of  
CC organ and tissue for transplantation e.g. of bone marrow, kidney, liver,  
CC pancreas, langerhans islet in pancreas and cornea, susceptibility  
CC diagnosis of genetic diseases and identifying individuals.  
XX  
SQ Sequence 17 BP; 3 A; 2 C; 9 G; 3 T; 0 other;  
SQ  
Query Match 100.0%; Score 11; DB 24; Length 17;  
Best Local Similarity 100.0%; Pred. No. 1.9e+03;  
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 tctcaccacc 11  
|||||  
DB 15 TCTCACCACACC 5  
RESULT 10  
AAS01746/C  
ID AAS01746 standard; DNA; 18 BP.  
XX  
AC AAS01746;  
XX  
DT 12-SEP-2001 (first entry)  
XX  
DE Glucanase genomic DNA sequencing primer Glucgene-1.  
XX  
KW Glucanase; endochitinase; exochitinase; cell-wall degradation; fungus;  
KW transgenic plant; plant pathogen; bacteria; seafood waste; shell; ss;  
KW chitin; chemical modification; glucan; sequencing primer; glucgene-1.  
XX  
OS Fusarium venenatum.  
XX  
PN WO200116353-A1.  
XX  
PD 08-MAR-2001.  
XX  
PF 30-AUG-2000; 2000WO-US23802.  
XX  
PR 30-AUG-1999; 99US-0151582.  
PR 11-AUG-2000; 2000US-0224946.  
PR 28-AUG-2000; 2000US-0649747.  
XX  
PA (NOVO ) NOVO NORDISK BIOTECH INC.  
PA (USDA ) US SEC OF AGRIC.  
PI Okubara PA, Blechl AE, Hohn TM, Berka RM;  
XX  
DR WPI; 2001-218524/22.  
XX  
PT Fusarium nucleic acids encoding polypeptides having glucanase,  
PT endochitinase or exochitinase activity, useful for producing transgenic  
PT plants which are resistant to plant pathogens, particularly Fusarium  
PT species -  
XX  
PS Disclosure; Page 80; 216pp; English.  
XX  
CC The sequence represents a sequencing primer for DNA encoding the Fusarium  
CC fungal enzyme, glucanase. Glucanase, endochitinase and exochitinase  
CC are polypeptides with cell-wall degrading activity, derived from Fusarium  
CC fungal genes. The associated nucleic acids can be used to produce  
CC transgenic plants which are resistant to plant pathogens, particularly  
CC Fusarium species. They can also be used to isolate homologous genes from  
CC fungi to obtain genes which protect host cells, including fungi, bacteria  
CC and plants against related fungal pathogens. The polypeptides, especially  
CC chitinases and glucanases, are useful for degrading seafood waste, such  
CC as shells that contain chitin, or for chemical modification of chitin or  
CC glucan.  
SQ Sequence 18 BP; 3 A; 2 C; 8 G; 5 T; 0 other;  
SQ  
Query Match 100.0%; Score 11; DB 22; Length 18;

Best Local Similarity 100.0%; Pred. No. 1.9e+03;  
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
OY 1 tctcaccac 11  
          |||||  
DB 17 TCTCACCAACC 7

RESULT 11  
ABL30757/C  
ID ABL30757 standard; DNA: 18 BP.

AC ABL30757;

DT 21-MAR-2002 (first entry)

XX Human HLA genotyping oligonucleotide SEQ ID NO 246.

XX Human; human leukocyte antigen; HLA; genotype; polymorphism;

XX immunogenetic; transplantation; genetic disease; ss.

OS Homo sapiens.

PI WO200192572-A1,

PD 06-DEC-2001.

PF 01-JUN-2001; 2001WO-JP04662.

PR 01-JUN-2000; 2000JP-0164798.

PS (NIST ) NISSHINBO IND INC.

PA (SYST-) SYSTEM RES INC.

PI Inoko H, Kagiya T, Ichihara T, Matsumura Y, Moriya S, Nishida M;

DR WPI; 2002-122074/16.

XX Human leukocyte antigen (HLA) typing, useful for judging HLA genotypes  
PT of individuals e.g. by determining immunogenetic differences when  
transplanting between them -

PS Claim 10; Page 139; 345pp; Japanese.

CC The invention relates to a typing kit for judging human leukocyte antigen  
CC (HLA) genotype of a sample by hybridising a substrate on which 10-24 base  
CC oligonucleotides (ABL30512-ABL31809) originating in the sequences of  
CC genes e.g. belonging to HLA class I antigens on human genome and  
CC containing gene polymorphisms as alloantigens have been immobilised as  
CC primers for amplification of cleaved nucleic acids relating to gene  
CC polymorphisms. The method is useful for judging HLA genotypes of  
CC individuals by determining immunogenetic differences before transplanting  
CC between them, providing genetic information to decide compatibility of  
CC organ and tissue for transplantation e.g. of bone marrow, kidney, liver,  
CC pancreas, Langerhans islet in pancreas and cornea, susceptibility  
CC diagnosis of genetic diseases and identifying individuals.

XX Sequence 18 BP; 3 A; 2 C; 9 G; 4 T; 0 other;

Query Match 100.0%; Score 11; DB 24; Length 18;  
Best Local Similarity 100.0%; Pred. No. 1.9e+03;  
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 tctcaccac 11

DB 15 TCTCACCAACC 5

RESULT 12  
ABL30790/C  
ID ABL30790 standard; DNA: 18 BP.

AC ABL30790;  
XX 21-MAR-2002 (first entry)

DT Human HLA genotyping oligonucleotide SEQ ID NO 279.

XX Human; human leukocyte antigen; HLA; genotype; polymorphism;

XX immunogenetic; transplantation; genetic disease; ss.

OS Homo sapiens.

PI WO200192572-A1.

PD 06-DEC-2001.

PF 01-JUN-2001; 2001WO-JP04662.

PR 01-JUN-2000; 2000JP-0164798.

PS (NIST ) NISSHINBO IND INC.

PA (SYST-) SYSTEM RES INC.

PI Inoko H, Kagiya T, Ichihara T, Matsumura Y, Moriya S, Nishida M;

DR WPI; 2002-122074/16.

XX Human leukocyte antigen (HLA) typing, useful for judging HLA genotypes  
PT of individuals e.g. by determining immunogenetic differences when  
transplanting between them -

PS Claim 10; Page 146; 345pp; Japanese.

CC The invention relates to a typing kit for judging human leukocyte antigen  
CC (HLA) genotype of a sample by hybridising a substrate on which 10-24 base  
CC oligonucleotides (ABL30512-ABL31809) originating in the sequences of  
CC genes e.g. belonging to HLA class I antigens on human genome and  
CC containing gene polymorphisms as alloantigens have been immobilised as  
CC primers for amplification of cleaved nucleic acids relating to gene  
CC polymorphisms. The method is useful for judging HLA genotypes of  
CC individuals by determining immunogenetic differences before transplanting  
CC between them, providing genetic information to decide compatibility of  
CC organ and tissue for transplantation e.g. of bone marrow, kidney, liver,  
CC pancreas, Langerhans islet in pancreas and cornea, susceptibility  
CC diagnosis of genetic diseases and identifying individuals.

XX Sequence 18 BP; 2 A; 2 C; 9 G; 5 T; 0 other;

Query Match 100.0%; Score 11; DB 24; Length 18;  
Best Local Similarity 100.0%; Pred. No. 1.9e+03;  
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 tctcaccac 11

DB 14 TCTCACCAACC 4

RESULT 13  
AAQ06439/C  
ID AAQ06439 standard; DNA: 19 BP.  
XX AAQ06439;  
DT 04-FEB-1991 (first entry)  
XX Oligonucleotide probe to a lambda HLA-DR-5.0 sequence associated with  
DE human type I diabetes mellitus.  
XX Insulin-dependent diabetes; systemic lupus erythematosus;  
KW Reiter's disease; ss.  
XX Homo sapiens.

PN US4965189-A.  
 XX  
 PD 23-OCT-1990.  
 XX  
 PF 01-JUL-1986; 86US-0880857.  
 XX  
 PR 01-JUL-1986; 86US-0880857.  
 XX  
 PA (UYMA-) UNIV MASSACHUSETTS.  
 XX  
 PI Owerbach D;  
 XX  
 DR WPI; 1990-341710/45.  
 XX  
 PT DQ beta gene oligo:nucleotide(s) - for detection of proclivity in  
 XX humans for development of type I diabetes mellitus  
 PS Disclosure; Col 3; 17pp; English.  
 XX  
 CC Probe may be used in tests for proclivity towards autoimmune  
 CC diseases such as insulin dependent diabetes, Reiter's disease etc.  
 CC Probes are highly specific, even able to differentiate between  
 CC restriction fragments of identical size, and may also be used in  
 CC tissue typing.  
 XX  
 SQ Sequence 19 BP; 2 A; 3 C; 9 G; 5 T; 0 other;

Query Match 100.0%; Score 11; DB 11; Length 19;  
 Best Local Similarity 100.0%; Pred. No. 1.9e+03;  
 Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 tctcaccacc 11  
 |||||  
 DB 14 TCTCACCACMC 4

RESULT 14  
 AAQ26219/C  
 ID AAQ26219 standard; DNA; 19 BP.

AC AAQ26219;

DT 04-JAN-1993 (first entry)

DE HLA-DR beta sub-type tailed probe DRB15 hybridising region.

XX Tissue typing; identity determination; disease susceptible; ss.

OS Synthetic.

PN WO9210589-A.

PD 25-JUN-1992.

PF 06-DEC-1991; 91WO-US09294.

PR 06-DEC-1990; 90US-0623098.

PA (HOFF) HOFFMANN LA ROCHE & CO AG F.

PI Apple RJ, Begovich AB, Bugawan T, Erlich HA, Griffith RL;

XX Scharf SJ;

DR WPI; 1992-234644/28.

PT Method for determining HLA-DR beta sub-type in DNA sample -

XX comprises amplification and hybridisation with probes and

PS primers, useful in tissue typing

CC Example; Page 40; 90pp; English.  
 CC The sequence is that of the hybridising region of tailed probe DRB15

CC for use in a method for determining HLA-DR beta sub-type in a nucleic  
 CC acid sample. The method allows specific nucleic acid sequences of the  
 CC second exon of HLA-DR beta genes to be amplified then probed for  
 CC identification of polymorphic sequences. The amplified DNA is useful for  
 CC typing homozygous or heterozygous samples from a variety of sources and  
 CC for detecting allelic variants not distinguishable by serological  
 CC methods. The typing system can be used in a reverse dot blot format which  
 CC is simple and rapid to perform, produces detectable signals in minutes  
 CC and can be utilised in tissue typing, determination of individual  
 CC identity and identifying disease susceptible individuals.  
 XX See also AAQ26092-Q26367.

SQ Sequence 19 BP; 3 A; 2 C; 9 G; 5 T; 0 other;

Query Match 100.0%; Score 11; DB 13; Length 19;  
 Best Local Similarity 100.0%; Pred. No. 1.9e+03;  
 Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 tctcaccacc 11  
 |||||  
 DB 15 TCTCACCACMC 5

RESULT 15  
 AAX79408/C  
 ID AAX79408 standard; DNA; 19 BP.

AC AAX79408;

DT 17-AUG-1999 (first entry)

DE HLA-DR typing probe VAL86.

XX Tissue typing; human leukocyte antigen; HLA; MHC; donor; allele; PCR;

KW major histocompatibility complex; bone marrow transplant; primer;

KW amplification; polymerase chain reaction; probe; polymorphism;

KW sequence-specific oligonucleotide probe hybridisation; ss.

OS Synthetic.

PN US5468611-A.

PD 21-NOV-1995.

PF 08-APR-1993; 93US-0045530.

PR 27-JUN-1990; 90US-0544218.

PA (BLOO-) BLOOD CENT RES FOUND INC.

PI Baxter-Lowe LA, Gorski JA;

DR WPI; 1996-010091/01.

PT Improved method for HLA typing - by DNA amplification and

XX sequence-specific oligo:nucleotide hybridisation, used to select

PS bone marrow donors

XX Disclosure; Column 19-20; 20pp; English.

CC A novel method of typing the human leukocyte antigen (HLA) of the major  
 CC histocompatibility complex (MHC), esp. for typing donors for bone marrow  
 CC transplants, involves determining if the donor tissue HLA-DR alleles are  
 CC selected from the gp.: HLA-DRW52C, DR12a,b, DR3a,n, DR3a-e, DRNew1,  
 CC DR6a, DR8a-d, DRW53a-c, DR4a-f, DR7, DR9, DR2a-c B3, DR2a-d B1, DR10 and  
 CC DR1a-c. The method uses PCR to amplify these regions followed by  
 CC sequence-specific oligonucleotide probe hybridisation (SSOPH) using the  
 CC probes AAX79365-X79429. SSOPH allows detection of polymorphisms that  
 CC predict differences at a single amino acid level thus reducing errors  
 CC and improving the chance of successfully matching tissues.

Sequence 19 BP; 2 A; 3 C; 9 G; 5 T; 0 other;

## Query Match

100.0%; Score 11; DB 17; Length 19;

Best Local Similarity 100.0%; Pred. No. 1.9e+03;

Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 tctaccaacc 11

Db 14 TCTCACCACC 4

Search completed: July 30, 2002, 00:01:20  
Job time: 4931 sec

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 29, 2002, 23:22:44 ; Search time 2542.47 Seconds  
(without alignments)  
58.395 Million cell updates/sec

Title: US-09-530-663B-17

Perfect score: 11

Sequence: 1 tctcaccacac 11

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 13736207 seqs, 6748477542 residues

Total number of hits satisfying chosen parameters: 27472414

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%  
Listing first 45 summaries

Database :

EST:\*  
1: em\_estbta:\*  
2: em\_esthum:\*  
3: em\_estli:\*  
4: em\_estmu:\*  
5: em\_estcov:\*  
6: em\_estropl:\*  
7: em\_estro:\*  
8: em\_hlc:\*  
9: gb\_estl:\*  
10: gb\_est2:\*  
11: gb\_hlc:\*  
12: gb\_gss:\*  
13: em\_gss\_hum:\*  
14: em\_gss\_inv:\*  
15: em\_gss\_pln:\*  
16: em\_gss\_vrt:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	11	100.0	36	12	AZ493433 1M0328A12
2	11	100.0	98	10	F30845 HSPD21538 H
3	11	100.0	100	9	A1922201 qm86e09.x
4	11	100.0	102	12	AZ233895 RPCI-23-8
5	11	100.0	108	9	AA184061 mo96f04.r
6	11	100.0	109	9	AA258662 zr61h08.r
7	11	100.0	111	10	BG953134 CM4-CT063
8	11	100.0	120	10	BF828693 MR2-HN003
9	11	100.0	122	10	BF085206 MR3-GN002
10	11	100.0	123	9	AA323418 EST28422
11	11	100.0	124	9	AA648082 ns10b07.r
12	11	100.0	124	10	BI040954 CM3-NT026
13	11	100.0	127	12	AZ753347 RPCI-24-1
14	11	100.0	129	12	AZ738998 RPCI-24-7
15	11	100.0	131	9	AT974770 T113239e
16	11	100.0	131	10	BI301600 UT-R-DL0
17	11	100.0	135	10	BI562481 603256389

C 18	11	100.0	136	9	AA052296 mb92g02.r
C 19	11	100.0	136	10	BE807681 ss29a01.y
C 20	11	100.0	137	10	BG342269 603274360
C 21	11	100.0	139	10	BI052985 RC0-GN027
C 22	11	100.0	145	9	BE033228 133585 MA
C 23	11	100.0	145	10	BE980472 UI-M-BG2-
C 24	11	100.0	146	9	AV028552 AV028552
C 25	11	100.0	146	12	A0997032 RPCI-23-3
C 26	11	100.0	148	10	BG954331 CM4-CT065
C 27	11	100.0	151	9	AA061917 m184a04.r
C 28	11	100.0	152	10	BE715956 MR2-HT075
C 29	11	100.0	154	10	BG133570 EST466558
C 30	11	100.0	155	9	AI137789 UI-R-E1-9
C 31	11	100.0	156	9	AV031373 AV031373
C 32	11	100.0	156	9	AV237319 AV237319
C 33	11	100.0	159	10	BF987685 QV0-GN014
C 34	11	100.0	159	10	BM052472 IC99b01.y
C 35	11	100.0	161	12	A2447050 1M0243D14
C 36	11	100.0	161	12	B07010 CSR1-8h5-u
C 37	11	100.0	162	9	BE231730 136547 MA
C 38	11	100.0	166	9	AA105067 mm68b10.r
C 39	11	100.0	167	10	BG190599 RST9672 A
C 40	11	100.0	170	9	AA576883 nm78a10.s
C 41	11	100.0	173	10	BF645790 NF028F04E
C 42	11	100.0	174	9	AA718915 ab45a11.s
C 43	11	100.0	174	9	AI705839 UI-R-AC1-
C 44	11	100.0	174	9	AM836760 QV1-LT003
C 45	11	100.0	174	9	BE031039 129375 MA

#### ALIGNMENTS

RESULT 1  
A2493433/c  
LOCUS  
DEFINITION  
1M0328A12F Mouse 10kb plasmid UUGC1M library Mus musculus genomic  
clone UUGC1M0328A12 F, DNA sequence.

ACCESSION  
A2493433  
A2493433.1 GI:10667114

VERSION  
KEYWORDS

SOURCE  
ORGANISM

REFERENCE  
AUTHORS

house mouse.  
Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sclerognathi; Muridae; Murinae; Mus.  
1 (bases 1 to 36)  
Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C.,  
Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly,  
M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A.  
and Wright,D., Weiss,R.  
Mouse whole genome scaffolding with paired end reads from 10kb  
plasmid inserts

JOURNAL  
COMMENT  
Unpublished (2000)  
Contact: Robert B. Weiss  
University of Utah Genome Center  
University of Utah

Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT  
84112, USA  
Tel: 801 585 5606  
Fax: 801 585 7177  
Email: ddunn@genetics.utah.edu  
Insert Length: 1000 Std Error: 0.00  
Plate: 0328 row: A column: 12  
Seq primer: CGTGTAAACGACGCCAGT  
Class: Plasmid ends  
High quality sequence stop: 36.  
Location/Qualifiers  
1..36

FEATURES  
source  
/organism="Mus musculus"  
/strain="C57BL/6J"  
/db\_xref="taxon:10090"  
/clone="UUGC1M0328A12"  
/clone\_lib="Mouse 10kb plasmid UUGC1M library"

```

/sex="Male"
/lab_host="E. Coli strain XL10-Gold, T1-resistant, F-"
/notes="Vector: PMD42uv; Purified genomic DNA from M.
musculalis C57BL/6J (male) was obtained from the Jackson
Laboratory Mouse DNA Resource
(http://www.jax.org/resources/documents/dnares/). The DNA
was hydrodynamically sheared by repeated passage through a
0.005 inch orifice at constant velocity. The sheared DNA
was blunt end-repaired with T4 DNA polymerase and T4
polynucleotide kinase. Adaptor oligonucleotides were
ligated to the blunt ends in high molar excess. The
adapted DNA was purified and size-selected for a 9.5 to
10.5 kb range using preparative agarose gel
electrophoresis. Vector DNA was prepared from a derivative
of pMD42 (GI4732114|gb|AF129072.1), a copy-number
inducible derivative of plasmid R1. The vector was ligated
with adaptors complementary to the insert adaptors and
purified. The sheared, adapted mouse DNA was annealed to
adapted vector DNA, and transformed into
chemically-competent E. coli XL10-Gold (Stratagene) cells
and selected for ampicillin resistance."
BASE COUNT      6 a      4 c      16 g      10 t
ORIGIN
Query Match      100.0%; Score 11; DB 12; Length 36;
Best Local Similarity 100.0%; Pred. No. 1.6e+04;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 tctcaccacc 11
Db 29 TCTCACCACC 19

RESULT 2
F30845/c 98 bp mRNA linear EST 13-MAY-1999
LOCUS HSPD21538 HM3 Homo sapiens cDNA clone s4000090D07, mRNA sequence.
DEFINITION F30845
VERSION F30845.1 GI:4816471
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 98)
Lanfranchi, G., Muraro, T., Caldara, F., Pacchioni, B., Pallavicini, A.,
Pandofo, D., Toppo, S., Trevisan, S., Scarso, S. and Valle, G.
Identification of 4370 expressed sequence tags from a
3'-end-specific cDNA library of human skeletal muscle by DNA
sequencing and filter hybridization
Genome Res. 6 (1), 35-42 (1996)
96276048
Contact: Valle G.
CRIBI Biotechnology Centre
University of Padua
Via Trieste 75, 35121 Padua, Italy
ABI Chromatograms and other information are available on WWW at
http://grip.bio.unipd.it.
FEATURES
Source
1. 98
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="s4000090D07"
/clone_lib="HM3"
/sex="Female"
/tissue_type="pectoral muscle (after mastectomy)"
/notes="Vector: pCDNAIL (Invitrogen); Site_1: BstXI;
Site_2: NotI. The library was constructed by G.
Lanfranchi. This library is not subcloned nor normalized.
The first strand cDNA was primed with a biotinylated
oligo-dT-NotI primer
(5'-biotin-AACCCGGCTCGAGCGCCGCTTTT-TTTT-TTTT-3'). The

```

```

ds cDNA was sonicated and size-selected in the range
350-550 bp. The 3' specific fragments were selected by
streptavidin coated magnetic beads, ligated to
non-palindromic BstXI adaptors, NotI digested and
directionally cloned into BstXI-NotI cut pCDNAIL vector."
BASE COUNT      42 a      11 c      21 g      24 t
ORIGIN
Query Match      100.0%; Score 11; DB 10; Length 98;
Best Local Similarity 100.0%; Pred. No. 2.1e+04;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 tctcaccacc 11
Db 67 TCTCACCACC 57

RESULT 3
A1292201 100 bp mRNA linear EST 30-NOV-1998
LOCUS qm86e09.x1 NCI-CGAP LUS Homo sapiens cDNA clone IMAGE:1895656 3'
DEFINITION A1292201
VERSION A1292201.1 GI:3934975
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 100)
NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
Unpublished (1997)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.
Emmert-Buck, M.D., Ph.D.
CDNA library Preparation: M. Bento Soares, Ph.D.
CDNA library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
www-bio.llnl.gov/bbrp/image/image.html
Seq primer: -40up from Gibco
High quality sequence stop: 1.
Location/Qualifiers
1. 100
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:1895656"
/clone_lib="NCI-CGAP_LUS"
/tissue_type="carcinoid"
/lab_host="DH10B"
/notes="Organ: lung; Vector: pT73D-Pac (Pharmacia) with a
modified polylinker; 1st strand cDNA was prepared from
neuroendocrine lung carcinoid, and was then primed with a
Not I - oligo(dT) primer. Double-stranded cDNA was ligated
to Eco RI adaptors (Pharmacia), digested with Not I and
cloned into the Not I and Eco RI sites of the modified
pT73 vector. Library is normalized. Library was
constructed by Bento Soares and M. Fatima Bonaldo. "
BASE COUNT      30 a      27 c      19 g      24 t
ORIGIN
Query Match      100.0%; Score 11; DB 9; Length 100;
Best Local Similarity 100.0%; Pred. No. 2.1e+04;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 tctcaccacc 11

```



DB	31	TCTCACCAACC	41		
RESULT	4				
LOCUS	A2233895				
DEFINITION	RPCT-23-82M10.TV	RPCT-23	Mus musculus genomic clone	RPCT-23-82M10,	
ACCESSION	A2233895				
VERSION	A2233895.1	GI:8541941			
KEYWORDS	GSS.				
SOURCE	house mouse.				
ORGANISM	Mus musculus				
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus. 1 (bases 1 to 102)				
AUTHORS	Zhao, S., Niemman, W., Feldblyum, T., Malek, J., Shatsman, S., Akinret, B., Levins, M., McGann, S., Tsegaye, G., Geer, K., Krol, M., de Jong, P. and Fraser, C.M.				
TITLE	Mouse BAC End Sequences from Library RPCT-23				
JOURNAL	Unpublished (1999)				
COMMENT	Other_GSSs: RPCT-23-82M10.TJ Contact: Shaying Zhao Department of Eukaryotic Genomics The Institute for Genomic Research 9712 Medical Center Dr., Rockville, MD 20850, USA Tel: 301 838 0200 Fax: 301 838 0208 Email: szhao@tigr.org Clones are derived from the mouse BAC library RPCT-23. For BAC library availability, please contact Pieter de Jong (pieter@eijong.med.buffalo.edu). Clones may be purchased from BACPAC Resources (http://bacpac.med.buffalo.edu/orderingframe.htm) or from Resea ch Genetics (info@resgen.com). BAC end page: http://www.tigr.org/tdb/bac_ends/mouse/BAC_end_intro.html Plate: 82 row: M column: 10 Seq primer: T7 Class: BAC ends.				
FEATURES					
source	Location/Qualifiers				
	1..102				
	/organism="Mus musculus"				
	/strain="C57BL/6J"				
	/db_xref="taxon:10090"				
	/clone="RPCT-23-82M10"				
	/clone_1lb="RPCT-23"				
	/sex="Female"				
	/lab_host="DH10B"				
	/note="Organ: Kidney/Brain; Vector: pBAC3.6; Site.1: EcoRI; Site.2: EcoRI; Female C57BL/6J mouse kidney and/or brain genomic DNA was isolated and partially digested with a combination of EcoRI and EcoRI Methylase. Size selected DNA was cloned into the pBAC3.6 vector at the EcoRI sites. The ligation products were transformed into DH10B electrocompetent cells (BRL Life Technologies)."				
BASE COUNT	37 a	18 c	17 g	30 t	
ORIGIN					
Query Match					
Best Local Similarity	100.0%;	Score 11;	DB 12;	Length 102;	
Matches	11;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;
0Y	1	tctcacgaacc	11		
Db	63	TCTCACCAACC	73		
RESULT	5				
LOCUS	AA184061				
DEFINITION	mouse6104.t1	Stratagene mouse testis (#937308)	Mus musculus	CDNA	
	clone IMAGE:567583	5', mRNA	sequence.		

ACCESSION	AA184061	GI:1767429	
VERSION	AA184061.1		
KEYWORDS	EST.		
SOURCE	house mouse.		
ORGANISM	Mus musculus		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.		
AUTHORS	1 (bases 1 to 108) Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T., Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M., Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B., Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and Waterston,R.		
TITLE	The WashU-HMI Mouse EST Project		
JOURNAL	Unpublished (1996)		
COMMENT	Contact: Marra M/Mouse EST Project WashU-HMI Mouse EST Project Washington University School of Medicine 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108 Tel: 314 286 1800 Fax: 314 286 1810 Email: mouseest@watson.wustl.edu This clone is available royalty-free through LNL; contact the IMAGE Consortium (info@image.llnl.gov) for further information. MG1:342231 Seq primer: -28m13 rev1 ET from Amersham High quality sequence stop: 96. Location/Qualifiers 1..108 /organism="Mus musculus" /strain="Inbred CD-1" /db_xref="taxon:10090" /clone_image="567583" /clone_lib="Stratagene mouse testis (#937308)" /sex="males" /tissue_type="testis" /dex_stage="10-12 week old" /lab_host="SOLR (kanamycin resistant)" /note="Organ: testis; Vector: pBluescript SK-; Site_1: EcoRI; Site_2: XhoI; Cloned unidirectionally. Primer: Oligo dT. Average insert size: 1.0 kb; Uni-ZAP XR Vector; ~5' adaptor sequence: 5' GAATTCGCACGCG 3' ~3' adaptor sequence: 5' CTCGAGCTTTT TTTT TTTT TTTT 3'."		
BASE COUNT	28 a 24 c 34 g 22 t		
ORIGIN			
Query Match	100.0%; Score 11; DB 9; Length 108;		
Best Local Similarity	100.0%; Pred. No. 2; Je+04;		
Matches 11; Conservative	0; Mismatches 0; Indels 0; Gaps 0;		
OY	1 tctcaccacac 11 		
Db	41 TCTCACCAC 31		
RESULT			
6 AA258662/c	109 bp mRNA linear EST 06-AUG-1997		
LOCUS	zrl1h08.r1 Soares_NHMPu_S1 Homo sapiens cDNA clone IMAGE:667935		
DEFINITION	5', mRNA sequence.		
ACCESSION	AA258662		
VERSION	AA258662.1 GI:1893786		
KEYWORDS	EST.		
SOURCE	human.		
ORGANISM	Homo sapiens		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.		
AUTHORS	1 (bases 1 to 109) Hillier,L., Allen,M., Bowles,L., Dubuque,T., Geisel,G., Jost,S., Kucaba,T., Lacy,M., Le,N., Lennon,G., Marra,M., Martin,J., Moore,B., Schellenberg,K., Steptoe,M., Tan,F., Theising,B., White,Y., Wylie,T., Waterston,R. and Wilson,R.		

TITLE WashU-Merck EST Project 1997  
JOURNAL Unpublished (1997)  
COMMENT Contact: Wilson RK  
Washington University School of Medicine  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
Tel: 314 286 1800  
Fax: 314 286 1810  
Email: est@wustl.wustl.edu  
This clone is available royalty-free through LNLN; contact the  
IMAGE Consortium (info@image.llnl.gov) for further information.  
Insert length: 1750 Std Error: 0.00  
Seq primer: -28m13 rev2 ET from Amersham.

FEATURES  
source  
1. 109  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:667935"  
/clone\_1ib="Soares,NHMPu,SI"  
/tissue\_type="Pooled human melanocyte, fetal heart, and  
pregnant uterus"  
/lab\_host="DH10B"  
/note="Organ: mixed (see below); Vector: pT73D-Pac  
(Pharmacia) with a modified polylinker; Site\_1: Not I;  
Site\_2: Eco RI; Equal amounts of plasmid DNA from three  
normalized libraries (melanocyte 2NDHM, pregnant uterus  
NBHPu, and fetal heart NBH19M) were mixed, and ss circles  
were made in vitro. Following HAP purification, this DNA  
was used as tracer in a subtractive hybridization  
reaction. The driver was PCR-amplified cDNAs from pools of  
5,000 clones made from the same 3 libraries. The pools  
consisted of I.M.A.G.E. clones 26032-26523,  
340488-345479, and 484488-489479."

BASE COUNT  
ORIGIN  
27 a 23 c 31 g 28 t

Query Match 100.0%; Score 11; DB 9; Length 109;  
Best Local Similarity 100.0%; Pred. No. 2.1e+04;  
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 tctaccacac 11  
|||||

Db 20 TCTCACCACAC 10

RESULT 7  
BG953134 111 bp mRNA linear EST 12-JUN-2001  
LOCUS CM4-CT0639-220101-695-d02 CT0639 Homo sapiens cDNA, mRNA sequence.  
DEFINITION BG953134  
ACCESSION BG953134  
VERSION BG953134.1 GI:14371305  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.  
1 (bases 1 to 111)  
Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,  
Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,  
Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H.,  
Brunstein,A., deoliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare  
,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and  
Simpson,A.J.  
Shotgun sequencing of the human transcriptome with ORF expressed  
sequence tags  
Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)

TITLE  
JOURNAL  
MEDLINE  
COMMENT Contact: Simpson A.J.G.  
Laboratory of Cancer Genetics  
Ludwig Institute for Cancer Research  
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,  
Brazil  
Tel: +55-11-2704922

Fax: +55-11-2707001  
Email: asimpson@ludwig.org.br  
This sequence was derived from the FAPESP/LICR Human Cancer Genome  
Project. This entry can be seen in the following URL  
(http://www.ludwig.org.br/scripts/gethtml2.pl?cl=CM4&c2=CM4-CT0639-  
220101-695-d02&c3=2001-01-22&t=1)  
Seq primer: puc 18 forward  
High quality sequence stop: 111.  
Location/Qualifiers  
1. 111  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone\_1ib="CT0639"  
/dev\_stage="Adult"  
/note="Organ: colon; Vector: puc18; Site\_1: SmaI; Site\_2:  
SmaI; A mini-library was made by cloning products derived  
from ORFESTS PCR (U.S. Letters Patent application No. 196  
,716 - Ludwig Institute for Cancer Research) profiles  
into the pUC 18 vector. Reverse transcription of tissue  
mRNA and cDNA amplification were performed under low  
stringency conditions."

BASE COUNT  
ORIGIN  
18 a 32 c 25 g 36 t

Query Match 100.0%; Score 11; DB 10; Length 111;  
Best Local Similarity 100.0%; Pred. No. 2.1e+04;  
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 tctaccacac 11  
|||||

Db 70 TCTCACCACAC 80

RESULT 8  
BF828693/c 120 bp mRNA linear EST 13-JAN-2001  
LOCUS BF828693  
DEFINITION MR2-HN0035-141200-015-e01 HN0035 Homo sapiens cDNA, mRNA sequence.  
ACCESSION BF828693  
VERSION BF828693.1 GI:12173470  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.  
1 (bases 1 to 120)  
Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,  
Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,  
Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H.,  
Brunstein,A., deoliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare  
,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and  
Simpson,A.J.  
Shotgun sequencing of the human transcriptome with ORF expressed  
sequence tags  
Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)

TITLE  
JOURNAL  
MEDLINE  
COMMENT Contact: Simpson A.J.G.  
Laboratory of Cancer Genetics  
Ludwig Institute for Cancer Research  
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,  
Brazil  
Tel: +55-11-2704922  
Fax: +55-11-2707001  
Email: asimpson@ludwig.org.br  
This sequence was derived from the FAPESP/LICR Human Cancer Genome  
Project. This entry can be seen in the following URL  
(http://www.ludwig.org.br/scripts/gethtml2.pl?cl=MR2&c2=MR2-HN0035-  
141200-015-e01&c3=2000-12-14&t=1)  
Seq primer: puc 18 forward  
High quality sequence stop: 110.  
Location/Qualifiers  
1. 120

BASE COUNT	48 a	21 c	32 g	21 t	performed under low stringency conditions."
ORIGIN					
Query Match	100.0%; Score 11; DB 10; Length 122;				
Best Local Similarity	Pred. No. 2,2e+04;				
Matches	11; Conservative	0; Mismatches	0; Indels	0; Gaps	0;
QY	1 tctcacaacc 11				
Db	39 TCTCACCACACC 29				
RESULT 10					
LOCUS	AA325418/c	123 bp	mRNA	linear	EST 20-APR-1997
DEFINITION	EST28422 Cerebellum II Homo sapiens cDNA 5' end, mRNA sequence.				
ACCESSION	AA325418				
VERSION	AA325418.1	GI:1977683			
KEYWORDS	EST.				
SOURCE	human.				
ORGANISM	Homo sapiens				
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.				
AUTHORS	1 (bases 1 to 123) Adams,M.D., Kerlavage,A.R., Fleischmann,R.D., Fuldner,R.A., Bult,C.J., Lee,N.J., Kirkness,E.F., Weinstock,K.G., Gocayne,J.D., White,O., Sutton,G., Blake,J.A., Brandon,R.C., Man'wal,C., Clayton,R.A., Cline,T.R., Cotton,M.D., Baile-Hughes,J., Geoghegan,N.S., Glodetz,A., L.M., Fitzhugh,W.M., Fritchman,J.L., Geoghegan,N.S., Glodetz,A., Grehm,C.L., Hanna,M.C., Hedblom,E., Hinkle,P.S., Jr., Kelley,J.M., Kelley,J.C., Liu,L.-I., Marmaros,S.M., Merrick,J.M., Moreno-Palmarques,R.F., McDonald,L.A., Nguyen,D.T., Pelligrino,S.M., Phillips,C.A., Ryder,S.E., Scott,J.L., Saudak,D.M., Shirley,R., Small,K.V., Spriggs,T.A., Uterback,T.R., Weidman,J.F., Li,Y., Bednarek,D.P., Cao,L., Cepeda,M.A., Coleman,T.A., Collins,E.J., Dinke,D., Feng,D.-F., Ferlie,A., Fischer,C., Hastings,G.A., He,W.M., Hu,J.S., Greene,J.M., Gruber,J., Hudson,P., Kim,A.K., Kozak,D.L., Kunsch,C., Hungjun,J., Li,H., Meissner,P.S., Olsen,H., Raymond,L., Wei,Y.F., Wang,J., Xu,C., Yu,G.L., Ruben,S.M., Dillion,P.J., Fannon,M.R., Rosen,C.A., Haseltine,W.A., Fields,C., Fraser,C.M. and Venter,J.C				
TITLE	Initial assessment of human gene diversity and expression patterns based upon 83 million nucleotides of cDNA sequence				
JOURNAL	Nature 377 (6547 Suppl), 3-174 (1995)				
MIDDLELINE	96026280				
COMMENT	Other ESTs: TNC168430 Contact: Kerlavage, AR Bioinformatics The Institute for Genomic Research 9712 Medical Center Drive, Rockville, MD 20850 USA Tel: 3018699056 Fax: 3018699423 Email: arkerlav@tigr.org For clone availability, additional sequence and expression information related to this EST, please check the TIGR Human Gene Index ( <a href="http://www.tigr.org/tdb/hgi/hgi.html">http://www.tigr.org/tdb/hgi/hgi.html</a> ) Seq primer: M13 Reverse. Location/Qualifiers 1..123 /organism="Homo sapiens" /db_xref="ATCC (inhost):1125942" /db_xref="taxon:9606" /clone_id="Cerebellum II" /tissue_type="Cerebellum" /dev_stage="adult" /note="Organ: brain; Vector: pBluescript SK-; Site_1: EcoRI; Site_2: XhoI"				
BASE COUNT	16 a	34 c	44 g	26 t	3 others
ORIGIN					

Query Match 100.0%; Score 11; DB 9; Length 123;  
 Best Local Similarity 100.0%; Pred. No. 2.2e+04;  
 Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 tctcaccacc 11  
 |||||

Db 61 TCTCACCACC 51

RESULT 11  
 AA648082/c 124 bp mRNA linear EST 29-OCT-1997  
 LOCUS ns10007.r1 NCI-CGAP\_Ew1 Homo sapiens cDNA clone IMAGE:1183189  
 DEFINITION similar to TR:5624778 G624778 E25.; mRNA sequence.  
 ACCESSION AA648082  
 VERSION AA648082.1 GI:2574511  
 KEYWORDS EST.  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 1 (bases 1 to 124)  
 NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.  
 National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
 Tumor Gene Index  
 Unpublished (1997)  
 Contact: Robert Strausberg, Ph.D.  
 Email: cgaps-r@mail.nih.gov  
 Tissue Procurement: Lee Helman, M.D., Michael R. Emmert-Buck, M.D.,  
 Ph.D.  
 cDNA Library Preparation: David B. Krizman, Ph.D.  
 cDNA Library Arrayed by: Greg Lennon, Ph.D.  
 DNA Sequencing by: Washington University Genome Sequencing Center  
 Clone distribution: NCI-CGAP clone distribution information can be  
 found through the I.M.A.G.E. Consortium/LLNL at:  
[www.bio.llnl.gov/dbp/image/image.html](http://www.bio.llnl.gov/dbp/image/image.html)

FEATURES  
 source  
 1..124  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone\_lib="NCI-CGAP\_Ew1"  
 /tissue\_type="Ewing's sarcoma"  
 /lab\_host="DH10B"  
 /note="Vector: PAMP10; mRNA made from Ewing's sarcoma,  
 cDNA made by oligo-dT priming. Non-directionally cloned.  
 Size-selected on agarose gel, average insert size 600 bp.  
 Reference: Krizman et al. (1996) Cancer Research  
 56:5380-5383."

BASE COUNT 32 a 18 c 45 g 29 t  
 ORIGIN

Query Match 100.0%; Score 11; DB 9; Length 124;  
 Best Local Similarity 100.0%; Pred. No. 2.2e+04;  
 Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 tctcaccacc 11  
 |||||

Db 62 TCTCACCACC 52

RESULT 12  
 BI040954 124 bp mRNA linear EST 14-JUN-2001  
 LOCUS CM3-NT0268-130201-748-d05\_1 NT0268 Homo sapiens cDNA, mRNA  
 DEFINITION sequence.  
 ACCESSION BI040954  
 VERSION BI040954.1 GI:14447580

KEYWORDS EST.  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 1 (bases 1 to 124)  
 Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,  
 Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,  
 Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H.,  
 Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare  
 M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and  
 Simpson,A.J.  
 Shotgun sequencing of the human transcriptome with ORF expressed  
 sequence tags  
 Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)

JOURNAL MEDLINE  
 COMMENT  
 Contact: Simpson A.J.G.  
 Laboratory of Cancer Genetics  
 Ludwig Institute for Cancer Research  
 Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,  
 Brazil  
 Tel: +55-11-2704922  
 Fax: +55-11-2707001  
 Email: [asimpson@ludwig.org.br](mailto:asimpson@ludwig.org.br)  
 This sequence was derived from the FAPESP/LICR Human Cancer Genome  
 Project. This entry can be seen in the following URL  
 (<http://www.ludwig.org.br/scripts/gethtml2.pl?tl=CM3&cl=CM3-NT0268-130201-748-d05-1&cl=2001-02-13&cl=4-1>)  
 Seq primer: puc 18 forward  
 High quality sequence stop: 119.  
 Location/Qualifiers  
 1..124  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone\_lib="NT0268"  
 /dev\_stage="Adult"  
 /note="Organ: nervous\_tumor; Vector: puc18; site\_1: SmaI;  
 site\_2: SmaI; A mini-library was made by cloning products  
 derived from ORESTES PCR (U.S. Letters Patent application  
 No. 196,716 - Ludwig Institute for Cancer Research)  
 profiles into the pUC 18 vector. Reverse transcription of  
 tissue mRNA and cDNA amplification were performed under  
 low stringency conditions."

BASE COUNT 32 a 23 c 32 g 37 t  
 ORIGIN

Query Match 100.0%; Score 11; DB 10; Length 124;  
 Best Local Similarity 100.0%; Pred. No. 2.2e+04;  
 Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 tctcaccacc 11  
 |||||

Db 88 TCTCACCACC 78

RESULT 13  
 A2753347 127 bp DNA linear GSS 25-JAN-2001  
 LOCUS RPCI-24-147P15.TJ RPCI-24 Mus musculus genomic clone RPCI-24-147P15  
 DEFINITION , DNA sequence.  
 ACCESSION A2753347  
 VERSION A2753347.1 GI:12538506  
 KEYWORDS GSS.  
 SOURCE house mouse.  
 ORGANISM Mus musculus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 1 (bases 1 to 127)  
 Zhao,S., Niernan,W., Malek,J., Shatsman,S., Akiret,B., Levins,M.,  
 Tsegaye,G., Geer,K., Krol,M., Shvartsbeyn,A., Gebregiorgis,E.,  
 Russell,D., de Jong,P. and Fraser,C.M.  
 Mouse BAC End sequences from Library RPCI-24

JOURNAL  
COMMENT

Unpublished (1999)  
Other-GSSS: RPCI-24-147P15.TV  
Contact: Shaying Zhao  
Department of Eukaryotic Genomics  
The Institute for Genomic Research  
9712 Medical Center Dr., Rockville, MD 20850, USA  
Tel: 301 838 0200  
Fax: 301 838 0208

Email: szhao@tigr.org  
Clones are derived from the mouse BAC library RPCI-24. For BAC library availability, please contact Pieter de Jong (pdejong@mail.cho.org). Clones may be purchased from BACPAC Resources (<http://www.choi.org/bacpac/orderingframe.htm>). BAC end page: [http://www.tigr.org/tdb/bac\\_ends/mouse/bac\\_end\\_intro.html](http://www.tigr.org/tdb/bac_ends/mouse/bac_end_intro.html)  
Plate: 147 row: P column: 15  
Seq primer: SP6  
Class: BAC ends.

FEATURES  
SOURCE

Location/Qualifiers  
1. .129  
/organism="Mus musculus"  
/strain="C57BL/6J"  
/db\_xref="taxon:10090"  
/clone="RPCI-24-147P15"  
/clone\_lib="RPCI-24"  
/sex="Male"  
/cell\_type="Spleen/Brain"  
/note="Vector: pTRABAC1; Site\_1: BamHI; Site\_2: BamHI; RPCI-24 Mouse BAC library produced by Pieter de Jong. The library was cloned in the pTRABAC1 cloning vector at the BamHI sites using MboI partially digested male C57BL/6J DNA."

BASE COUNT 42 a 42 c 24 g 19 t  
ORIGIN

Query Match 100.0%; Score 11; DB 12; Length 127;  
Best Local Similarity 100.0%; Pred. No. 2.2e+04;

Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 tctcaccacc 11  
|||||  
Db 59 TCTCACCACACC 69

## RESULT 14

A2738998

LOCUS A2738998 129 bp DNA linear GSS 25-JAN-2001  
DEFINITION RPCI-24-71F6.TV RPCI-24 Mus musculus genomic clone RPCI-24-71F6,  
DNA sequence.

ACCESSION A2738998  
VERSION A2738998.1 GI:12508671

KEYWORDS GSS.  
SOURCE house mouse.  
ORGANISM Mus musculus

## REFERENCE

AUTHORS

1 (bases 1 to 129)  
Zhao, S., Nierman, W., Malek, J., Shatsman, S., Akinet, B., Levins, M., Russell, D., de Jong, P. and Fraser, C.M.  
Mouse BAC End Sequences from Library RPCI-24  
Unpublished (1999)  
Other-GSSS: RPCI-24-71F6.TV  
Contact: Shaying Zhao  
Department of Eukaryotic Genomics  
The Institute for Genomic Research  
9712 Medical Center Dr., Rockville, MD 20850, USA  
Tel: 301 838 0200  
Fax: 301 838 0208  
Email: szhao@tigr.org

JOURNAL  
COMMENT

Clones are derived from the mouse BAC library RPCI-24. For BAC library availability, please contact Pieter de Jong (pdejong@mail.cho.org). Clones may be purchased from BACPAC

FEATURES  
SOURCE

Resources (<http://www.choi.org/bacpac/orderingframe.htm>). BAC end page: [http://www.tigr.org/tdb/bac\\_ends/mouse/bac\\_end\\_intro.html](http://www.tigr.org/tdb/bac_ends/mouse/bac_end_intro.html)  
Plate: 71 row: F column: 6  
Seq primer: T7  
Class: BAC ends.

Location/Qualifiers  
1. .129  
/organism="Mus musculus"  
/strain="C57BL/6J"  
/db\_xref="taxon:10090"  
/clone="RPCI-24-71F6"  
/clone\_lib="RPCI-24"  
/sex="Male"  
/cell\_type="Spleen/Brain"  
/note="Vector: pTRABAC1; Site\_1: BamHI; Site\_2: BamHI; RPCI-24 Mouse BAC library produced by Pieter de Jong. The library was cloned in the pTRABAC1 cloning vector at the BamHI sites using MboI partially digested male C57BL/6J DNA."

BASE COUNT 25 a 19 c 21 g 64 t  
ORIGIN

Query Match 100.0%; Score 11; DB 12; Length 129;  
Best Local Similarity 100.0%; Pred. No. 2.2e+04;

Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 tctcaccacc 11  
|||||  
Db 89 TCTCACCACACC 99

## RESULT 15

A1974770/c

LOCUS A1974770 131 bp mRNA linear EST 26-AUG-1999  
DEFINITION T113239e KV2 Medicago truncatula cDNA clone pKV2-1K8, mRNA  
sequence.

ACCESSION A1974770  
VERSION A1974770.1 GI:5777151

KEYWORDS EST.  
SOURCE barrel medic.  
ORGANISM Medicago truncatula

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Trifoliales; Medicago.

## REFERENCE

AUTHORS

1 (bases 1 to 131)  
Vandenbosch, K., Hur, J., Moore, J., Beremand, P., Peng, H. and Ellis, L.  
ESTs from Sinorhizobium-inoculated roots of Medicago truncatula (1999b)

Unpublished (1999)  
Contact: Vandenbosch K  
Department of Biology  
Texas A&M University  
College Station, TX 77843-3258, USA  
Tel: 409 845 7707  
Fax: 409 845 2891

JOURNAL  
COMMENT

Email: kate@mail.bio.tamu.edu  
Other name: 02-KV2-3F4; date: 8/5/99; Submitted to the Database of Expressed Sequence Tags (dbEST) on 08/25/99; More information is available at '<http://chrystle.tamu.edu/medicago/>'.  
Seq primer: SKmod (CTA GAA CTA GTG GAT CC).

FEATURES  
SOURCE

Location/Qualifiers  
1. .131  
/organism="Medicago truncatula"  
/cultivar="genotype A17"  
/db\_xref="taxon:3880"  
/clone="pKV2-1K8"  
/clone\_lib="KV2"  
/tissue\_type="Seedling roots"  
/dev\_stage="2 days post-inoculation with Sinorhizobium meliloti"  
/lab\_host="E. coli strain SOLR"

/note="Vector: Bluescript SK -; Site\_1: EcoRI; Site\_2:  
XhoI; cDNA was prepared from polyA+ enriched RNA. The  
cDNA was directionally ligated into the Unizap XR vector  
from Stratagene and packaged using Gigapack III Gold  
packaging extracts. Plasmids containing cDNA inserts  
were excised from the recombinant lambda-Zap phage using  
Ex-assist helper phage and propagated in SOLR cells."

BASE COUNT 22 a 13 c 43 g 53 t  
ORIGIN

## Query Match

Best Local Similarity 100.0%; Score 11; DB 9; Length 131;  
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ttctaccacc 11  
Db 112 TCTCACCAACC 102

Search completed: July 29, 2002, 23:22:49  
Job time: 6825 sec

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 29, 2002, 23:56:14 ; Search time 65.09 Seconds  
(without alignments)  
41.511 Million cell updates/sec

Title: US-09-530-663B-17

Perfect score: 11

Sequence: 1 tctcaccacac 11

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 383533 seqs, 122816752 residues

Total number of hits satisfying chosen parameters: 767066

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued\_Patents\_NA:\*

1: /cgn2\_6/ptodata/2/lna/5A.COMB.seq:\*  
2: /cgn2\_6/ptodata/2/lna/5B.COMB.seq:\*  
3: /cgn2\_6/ptodata/2/lna/6A.COMB.seq:\*  
4: /cgn2\_6/ptodata/2/lna/6B.COMB.seq:\*  
5: /cgn2\_6/ptodata/2/lna/PCRN.COMB.seq:\*  
6: /cgn2\_6/ptodata/2/lna/backfiles1.seq:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length DB	ID	Description
1	11	100.0	12	1	US-08-025-038-16 Sequence 16, Appl
2	11	100.0	16	1	US-08-050-073-77 Sequence 77, Appl
3	11	100.0	16	1	US-08-050-073-168 Sequence 168, App
4	11	100.0	17	1	US-08-050-073-223 Sequence 223, App
5	11	100.0	19	1	US-08-050-073-191 Sequence 191, App
6	11	100.0	25	4	US-09-441-346A-10 Sequence 10, Appl
7	11	100.0	54	1	US-08-758-306-1354 Sequence 1354, Ap
8	11	100.0	69	3	US-08-463-903-76 Sequence 76, Appl
9	11	100.0	69	4	US-07-935-695-76 Sequence 76, Appl
10	11	100.0	74	3	US-08-463-903-77 Sequence 77, Appl
11	11	100.0	74	4	US-07-935-695-77 Sequence 77, Appl
12	11	100.0	220	2	US-08-933-616-2 Sequence 2, Appl
13	11	100.0	268	1	US-08-039-137-16 Sequence 16, Appl
14	11	100.0	268	1	US-08-025-038-32 Sequence 32, Appl
15	11	100.0	269	1	US-08-025-038-33 Sequence 33, Appl
16	11	100.0	269	1	US-08-025-038-34 Sequence 34, Appl
17	11	100.0	269	1	US-08-025-038-37 Sequence 37, Appl
18	11	100.0	269	1	US-08-050-073-1 Sequence 1, Appl
19	11	100.0	269	1	US-08-050-073-3 Sequence 3, Appl
20	11	100.0	269	1	US-08-050-073-5 Sequence 5, Appl
21	11	100.0	269	1	US-08-050-073-7 Sequence 7, Appl
22	11	100.0	269	1	US-08-050-073-11 Sequence 11, Appl
23	11	100.0	269	1	US-08-050-073-13 Sequence 13, Appl
24	11	100.0	269	1	US-08-050-073-14 Sequence 14, Appl
25	11	100.0	269	1	US-08-050-073-15 Sequence 15, Appl
26	11	100.0	269	1	US-08-050-073-18 Sequence 18, Appl
27	11	100.0	269	1	US-08-050-073-19 Sequence 19, Appl

C 28	11	100.0	269	1	US-08-050-073-20	Sequence 20, Appl
C 29	11	100.0	269	1	US-08-050-073-21	Sequence 21, Appl
C 30	11	100.0	269	1	US-08-050-073-23	Sequence 23, Appl
C 31	11	100.0	269	1	US-08-050-073-24	Sequence 24, Appl
C 32	11	100.0	269	1	US-08-050-073-25	Sequence 25, Appl
C 33	11	100.0	269	1	US-08-050-073-29	Sequence 29, Appl
C 34	11	100.0	269	1	US-08-050-073-33	Sequence 33, Appl
C 35	11	100.0	269	1	US-08-050-073-34	Sequence 34, Appl
C 36	11	100.0	269	1	US-08-050-073-36	Sequence 36, Appl
C 37	11	100.0	269	1	US-08-050-073-38	Sequence 38, Appl
C 38	11	100.0	269	1	US-08-050-073-39	Sequence 39, Appl
C 39	11	100.0	269	1	US-08-050-073-43	Sequence 43, Appl
C 40	11	100.0	269	1	US-08-050-073-45	Sequence 45, Appl
C 41	11	100.0	269	1	US-08-050-073-46	Sequence 46, Appl
C 42	11	100.0	269	1	US-08-050-073-48	Sequence 48, Appl
C 43	11	100.0	269	1	US-08-050-073-52	Sequence 52, Appl
C 44	11	100.0	269	1	US-08-050-073-53	Sequence 53, Appl
C 45	11	100.0	269	1	US-08-050-073-69	Sequence 69, Appl

## ALIGNMENTS

RESULT 1  
US-08-025-038-16  
; Sequence 16, Application US/08025038  
; Patent No. 5545526  
; GENERAL INFORMATION:  
; APPLICANT: BAXTER-LOWE, Lee-Ann  
; TITLE OF INVENTION: Method For HLA Typing  
; NUMBER OF SEQUENCES: 46  
; CORRESPONDENCE ADDRESSES:  
; ADDRESSEE: Foley & Lardner  
; STREET: 777 E. Wisconsin Avenue  
; CITY: Milwaukee  
; STATE: Wisconsin  
; COUNTRY: USA  
; ZIP: 53202-5367  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/025,038  
; FILING DATE: 19930301  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 07/544,218  
; FILING DATE: 27-JUN-1990  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Meyers, Philip G.  
; REGISTRATION NUMBER: 30,478  
; REFERENCE/DOCKET NUMBER: 204 854  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (414)289-3761  
; TELEFAX: (414)289-3791  
; INFORMATION FOR SEQ ID NO: 16:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 12 base pairs  
; TYPE: NUCLEIC ACID  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; US-08-025-038-16

Query Match 100.0%, Score 11; DB 1; Length 12;  
Best Local Similarity 100.0%; Pred. No. 2.7e+02;  
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 tctcaccac 11  
Db 2 tctcaccac 12

```
RESULT 2
US-08-050-073-77/c
; Sequence 77, Application US/08050073
; Patent No. 5567809
; GENERAL INFORMATION:
; APPLICANT: Apple, Raymond J.
; APPLICANT: Begovich, Ann B.
; APPLICANT: Bugawan, Teodorica L.
; APPLICANT: Erlich, Henry A.
; APPLICANT: Griffith, Robert L.
; APPLICANT: Scharf, Stephen J.
; TITLE OF INVENTION: Methods and Reagents for HLA DRbeta DNA
; TITLE OF INVENTION: Typing
; NUMBER OF SEQUENCES: 315
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Hoffmann-La Roche Inc.
; STREET: 340 Kingsland Street
; CITY: Nutley
; STATE: New Jersey
; COUNTRY: U.S.A.
; ZIP: 07110
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/050,073
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Petry, Douglas A.
; REGISTRATION NUMBER: 35,321
; REFERENCE/DOCKET NUMBER: 8769
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (510) 814-2974
; TELEFAX: (510) 814-2977
; INFORMATION FOR SEQ ID NO: 77:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 16 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: genomic DNA
; US-08-050-073-77

Query Match 100.0%; Score 11; DB 1; Length 16;
Best Local Similarity 100.0%; Pred. No. 2.7e+02;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 tctaccacac 11
|||||
DB 13 TCTCACCAC 3

RESULT 3
US-08-050-073-168/c
; Sequence 168, Application US/08050073
; Patent No. 5567809
; GENERAL INFORMATION:
; APPLICANT: Apple, Raymond J.
; APPLICANT: Begovich, Ann B.
; APPLICANT: Bugawan, Teodorica L.
; APPLICANT: Erlich, Henry A.
; APPLICANT: Griffith, Robert L.
; APPLICANT: Scharf, Stephen J.
; TITLE OF INVENTION: Methods and Reagents for HLA DRbeta DNA
; TITLE OF INVENTION: Typing
; NUMBER OF SEQUENCES: 315
; CORRESPONDENCE ADDRESS:
```

```
; ADDRESSEE: Hoffmann-La Roche Inc.
; STREET: 340 Kingsland Street
; CITY: Nutley
; STATE: New Jersey
; COUNTRY: U.S.A.
; ZIP: 07110
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/050,073
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Petry, Douglas A.
; REGISTRATION NUMBER: 35,321
; REFERENCE/DOCKET NUMBER: 8769
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (510) 814-2974
; TELEFAX: (510) 814-2977
; INFORMATION FOR SEQ ID NO: 168:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 16 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: genomic DNA
; US-08-050-073-168
```

```
Query Match 100.0%; Score 11; DB 1; Length 16;
Best Local Similarity 100.0%; Pred. No. 2.7e+02;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 tctaccacac 11
|||||
DB 13 TCTCACCAC 3

RESULT 4
US-08-050-073-223/c
; Sequence 223, Application US/08050073
; Patent No. 5567809
; GENERAL INFORMATION:
; APPLICANT: Apple, Raymond J.
; APPLICANT: Begovich, Ann B.
; APPLICANT: Bugawan, Teodorica L.
; APPLICANT: Erlich, Henry A.
; APPLICANT: Griffith, Robert L.
; APPLICANT: Scharf, Stephen J.
; TITLE OF INVENTION: Methods and Reagents for HLA DRbeta DNA
; TITLE OF INVENTION: Typing
; NUMBER OF SEQUENCES: 315
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Hoffmann-La Roche Inc.
; STREET: 340 Kingsland Street
; CITY: Nutley
; STATE: New Jersey
; COUNTRY: U.S.A.
; ZIP: 07110
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/050,073
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Petry, Douglas A.
```



REGISTRATION NUMBER: 35,321  
REFERENCE/DOCKET NUMBER: 8769  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (510) 814-2974  
TELEFAX: (510) 814-2977  
INFORMATION FOR SEQ ID NO: 223:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 17 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: genomic DNA  
US-08-050-073-223

Query Match 100.0%; Score 11; DB 1; Length 17;  
Best Local Similarity 100.0%; Pred. No. 2.7e+02;  
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1 tctaccacc 11  
|||||  
Db 14 TCTCACCACC 4

RESULT 5  
US-08-050-073-191/C  
Sequence 191, Application US/08050073  
Patent No. 5567809  
GENERAL INFORMATION:  
APPLICANT: Apple, Raymond J.  
APPLICANT: Begovich, Ann B.  
APPLICANT: Bugawan, Teodorica L.  
APPLICANT: Erlich, Henry A.  
APPLICANT: Griffith, Robert L.  
APPLICANT: Scharf, Stephen J.  
TITLE OF INVENTION: Methods and Reagents for HLA DRbeta DNA  
TITLE OF INVENTION: Typing  
NUMBER OF SEQUENCES: 315  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Hoffmann-La Roche Inc.  
STREET: 340 Kingsland Street  
CITY: Nutley  
STATE: New Jersey  
COUNTRY: U.S.A.  
ZIP: 07110  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/050,073  
FILING DATE:  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Petry, Douglas A.  
REGISTRATION NUMBER: 35,321  
REFERENCE/DOCKET NUMBER: 8769  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (510) 814-2974  
TELEFAX: (510) 814-2977  
INFORMATION FOR SEQ ID NO: 191:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 19 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: genomic DNA  
US-08-050-073-191

Query Match 100.0%; Score 11; DB 1; Length 19;  
Best Local Similarity 100.0%; Pred. No. 2.7e+02;

Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Oy 1 tctaccacc 11  
|||||  
Db 15 TCTCACCACC 5

RESULT 6  
US-09-441-346A-10/C  
Sequence 10, Application US/09441346A  
Patent No. 624258  
GENERAL INFORMATION:  
APPLICANT: Sheppard, Paul O.  
APPLICANT: Piddington, Christopher S.  
APPLICANT: Ellsworth, Jeff L.  
TITLE OF INVENTION: TESTIS-SPECIFIC GLYCOPROTEIN ZPEP10  
FILE REFERENCE: 98-34  
CURRENT APPLICATION NUMBER: US/09/441,346A  
CURRENT FILING DATE: 1999-11-16  
PRIOR APPLICATION NUMBER: 60/109,216  
PRIOR FILING DATE: 1998-11-20  
NUMBER OF SEQ ID NOS: 14  
SOFTWARE: FastSeq for Windows Version 3.0  
SEQ ID NO 10  
LENGTH: 25  
TYPE: DNA  
ORGANISM: Artificial Sequence  
FEATURE:  
OTHER INFORMATION: Oligonucleotide ZC16,187  
US-09-441-346A-10

Query Match 100.0%; Score 11; DB 4; Length 25;  
Best Local Similarity 100.0%; Pred. No. 2.8e+02;  
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1 tctaccacc 11  
|||||  
Db 22 TCTCACCACC 12

RESULT 7  
US-08-758-306-1354/C  
Sequence 1354, Application US/08758306  
Patent No. 5807743  
GENERAL INFORMATION:  
APPLICANT: Scinichomb, Dan T.  
APPLICANT: McSwigen, James A.  
TITLE OF INVENTION: METHOD AND REAGENT FOR THE  
TITLE OF INVENTION: TREATMENT OF DISEASES  
TITLE OF INVENTION: ASSOCIATED WITH  
TITLE OF INVENTION: INTERLEUKIN-2 RECEPTOR  
NUMBER OF SEQUENCES: 1379  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Lyon & Lyon  
STREET: 633 West Fifth Street  
CITY: Los Angeles  
STATE: California  
COUNTRY: U.S.A.  
ZIP: 90071-2066  
COMPUTER READABLE FORM:  
MEDIUM TYPE: 3.5" Diskette, 1.44 Mb  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: IBM P.C. DOS 5.0  
SOFTWARE: FastSeq Version 1.5  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/758,306  
FILING DATE: December 3, 1996  
CLASSIFICATION: 514  
PRIOR APPLICATION DATA:

```
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Warburg, Richard J.
; REGISTRATION NUMBER: 32,327
; REFERENCE/DOCKET NUMBER: 212/132
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (213) 489-1600
; TELEFAX: (213) 955-0440
; TELEX: 67-3510
; INFORMATION FOR SEQ ID NO: 1354:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 54 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; US-08-758-306-1354
```

```
Query Match 100.0%; Score 11; DB 1; Length 54;
Best Local Similarity 100.0%; Pred. No. 2.9e+02;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 1 tctcaccacc 11
|||
Db 11 TCTCACCAACC 1
```

```
RESULT 8
US-08-463-903-76/c
; Sequence 76, Application US/08463903
; Patent No. 6071515
; GENERAL INFORMATION:
; APPLICANT: Mezes, Peter S.
; APPLICANT: Richard, Ruth A.
; APPLICANT: Afholter, Joseph A.
; APPLICANT: Kotite, Nicolas J.
; TITLE OF INVENTION: Dimer and Multimer Forms of Single Chain Polypeptides
; FILE REFERENCE: 40224A US
; CURRENT APPLICATION NUMBER: US/08/463,903
; EARLIER FILING DATE: 1995-06-05
; EARLIER APPLICATION NUMBER: US 07/935,695
; NUMBER OF SEQ ID NOS: 102
; SOFTWARE: MS-Word for Windows, Ver. 7.0
; SEQ ID NO 76
; LENGTH: 69
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; NAME/KEY: 0101F primer
; LOCATION: 1..69
; US-08-463-903-76
```

```
Query Match 100.0%; Score 11; DB 3; Length 69;
Best Local Similarity 100.0%; Pred. No. 2.9e+02;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 1 tctcaccacc 11
|||
Db 41 TCTCACCAACC 31
```

```
RESULT 9
US-07-935-695-76/c
; Sequence 76, Application US/07935695
; Patent No. 6329507
; GENERAL INFORMATION:
; APPLICANT: Mezes, Peter S.
; APPLICANT: Richard, Ruth A.
; APPLICANT: Afholter, Joseph A.
; APPLICANT: Kotite, Nicolas J.
```

```
; TITLE OF INVENTION: Dimer and Multimer Forms of Single Chain Polypeptides
; FILE REFERENCE: 40224A US
; CURRENT APPLICATION NUMBER: US/07/935,695
; CURRENT FILING DATE: 1992-08-21
; PRIOR APPLICATION NUMBER: US 08/463,903
; PRIOR FILING DATE: 1995-06-05
; NUMBER OF SEQ ID NOS: 102
; SOFTWARE: MS-Word for Windows, Ver. 7.0
; SEQ ID NO 76
; LENGTH: 69
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; NAME/KEY: 0101F primer
; LOCATION: 1..69
; OTHER INFORMATION:
; US-07-935-695-76
```

```
Query Match 100.0%; Score 11; DB 4; Length 69;
Best Local Similarity 100.0%; Pred. No. 2.9e+02;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 1 tctcaccacc 11
|||
Db 41 TCTCACCAACC 31
```

```
RESULT 10
US-08-463-903-77
; Sequence 77, Application US/08463903
; Patent No. 6071515
; GENERAL INFORMATION:
; APPLICANT: Mezes, Peter S.
; APPLICANT: Richard, Ruth A.
; APPLICANT: Afholter, Joseph A.
; APPLICANT: Kotite, Nicolas J.
; TITLE OF INVENTION: Dimer and Multimer Forms of Single Chain Polypeptides
; FILE REFERENCE: 40224A US
; CURRENT APPLICATION NUMBER: US/08/463,903
; EARLIER FILING DATE: 1995-06-05
; EARLIER APPLICATION NUMBER: US 07/935,695
; NUMBER OF SEQ ID NOS: 102
; SOFTWARE: MS-Word for Windows, Ver. 7.0
; SEQ ID NO 77
; LENGTH: 74
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; NAME/KEY: 0101R primer
; LOCATION: 1..74
; US-08-463-903-77
```

```
Query Match 100.0%; Score 11; DB 3; Length 74;
Best Local Similarity 100.0%; Pred. No. 3e+02;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 1 tctcaccacc 11
|||
Db 32 tctcaccacc 42
```

```
RESULT 11
US-07-935-695-77
; Sequence 77, Application US/07935695
; Patent No. 6329507
; GENERAL INFORMATION:
; APPLICANT: Mezes, Peter S.
; APPLICANT: Richard, Ruth A.
; APPLICANT: Afholter, Joseph A.
; APPLICANT: Kotite, Nicolas J.
```

;; TITLE OF INVENTION: Dimer and Multimer Forms of Single Chain Polypeptides  
;; FILE REFERENCE: 40224A US  
;; CURRENT APPLICATION NUMBER: US/07/935,695  
;; CURRENT FILING DATE: 1992-08-21  
;; PRIOR APPLICATION NUMBER: US 08/463,903  
;; PRIOR FILING DATE: 1995-06-05  
;; NUMBER OF SEQ ID NOS: 102  
;; SOFTWARE: MS-Word for Windows, Ver. 7.0  
;; SEQ ID NO 77  
;; LENGTH: 74  
;; TYPE: DNA  
;; ORGANISM: Artificial Sequence  
;; FEATURE:  
;; NAME/KEY: 0101r primer  
;; LOCATION: 1..74  
;; OTHER INFORMATION: :  
US-07-935-695-77

Query Match 100.0%; Score 11; DB 4; Length 74;  
Best Local Similarity 100.0%; Pred. No. 3e+02;  
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 tctcaccacc 11  
|||||  
Db 32 tctcaccacc 42

RESULT 12  
US-08-933-616-2  
; Sequence 2, Application US/08933616  
; Patent No. 5869331  
; GENERAL INFORMATION:  
; APPLICANT: Dornburg, Ralph C.  
; TITLE OF INVENTION: Cell-Type Specific Gene Transfer Using  
; TITLE OF INVENTION: Retroviral Vectors Containing Antibody-Envelope Fusion  
; NUMBER OF SEQUENCES: 5  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Richard R. Muccino  
; STREET: P.O. Box 1267  
; CITY: Princeton  
; STATE: New Jersey  
; COUNTRY: USA  
; ZIP: 08551  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentin Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/933,616  
; FILING DATE:  
; CLASSIFICATION:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US/08/205,980  
; FILING DATE: 04-MAR-1994  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Muccino, Richard R.  
; REGISTRATION NUMBER: 32,538  
; REFERENCE/DOCKET NUMBER: UMD1-025  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (609) 466-3407  
; TELEFAX: (609) 466-2760  
; INFORMATION FOR SEQ ID NO: 2:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 220 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: unknown  
; TOPOLOGY: unknown  
; MOLECULE TYPE: DNA (genomic)  
US-08-933-616-2

Query Match 100.0%; Score 11; DB 2; Length 220;  
Best Local Similarity 100.0%; Pred. No. 3.2e+02;  
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 tctcaccacc 11  
|||||  
Db 93 TCTCACCACC 103

RESULT 13  
US-08-039-137-16/c  
; Sequence 16, Application US/08039137  
; Patent No. 5759771  
; GENERAL INFORMATION:  
; APPLICANT: Tilius J.G., Marcel  
; TITLE OF INVENTION: Method of Determining a Genotype by  
; TITLE OF INVENTION: Comparing the Nucleotide Sequence of Members of a Gene  
; Patent No. 5759771  
; TITLE OF INVENTION: Family and Kit Therefor  
; NUMBER OF SEQUENCES: 44  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Dehlinger & Associates  
; STREET: 350 Cambridge Avenue, Suite 250  
; CITY: Palo Alto  
; STATE: CA  
; COUNTRY: USA  
; ZIP: 94306  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentin Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/039,137  
; FILING DATE: 14-APR-1993  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Fabian, Gary R.  
; REGISTRATION NUMBER: 33,875  
; REFERENCE/DOCKET NUMBER: 0550-0024.10  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (415) 324-0880  
; TELEFAX: (415) 324-0960  
; INFORMATION FOR SEQ ID NO: 16:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 268 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: double  
; TOPOLOGY: linear  
; MOLECULE TYPE: cDNA to mRNA  
; HYPOTHETICAL: NO  
; ANTI-SENSE: NO  
; INDIVIDUAL SOURCE:  
; INDIVIDUAL ISOLATE: DRI cDNA, Fig. 1  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: 1..267  
US-08-039-137-16

Query Match 100.0%; Score 11; DB 1; Length 268;  
Best Local Similarity 100.0%; Pred. No. 3.2e+02;  
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 tctcaccacc 11  
|||||  
Db 247 TCTCACCACC 237

RESULT 14  
US-08-025-038-32/c  
; Sequence 32, Application US/08025038

```
; Patent No. 5545526
; GENERAL INFORMATION:
; APPLICANT: BAXTER-LOWE, Lee-Ann
; TITLE OF INVENTION: Method For HLA Typing
; NUMBER OF SEQUENCES: 46
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Foley & Lardner
; STREET: 777 E. Wisconsin Avenue
; CITY: Milwaukee
; STATE: Wisconsin
; COUNTRY: USA
; ZIP: 53202-5367
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/025,038
; FILING DATE: 19930301
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/544,218
; FILING DATE: 27-JUN-1990
; ATTORNEY/AGENT INFORMATION:
; NAME: Meyers, Philip G.
; REGISTRATION NUMBER: 30,478
; REFERENCE/DOCKET NUMBER: 204 854
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (414)289-3761
; TELEFAX: (414)289-3791
; INFORMATION FOR SEQ ID NO: 32:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 269 base pairs
; TYPE: NUCLEIC ACID
; STRANDEDNESS: single
; TOPOLOGY: linear
;
US-08-025-038-32

Query Match      100.0%; Score 11; DB 1; Length 269;
Best Local Similarity 100.0%; Pred. No. 3.2e+02;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 tctcaccacc 11
        |||
Db      249 TCTCACCAC 239

RESULT 15
US-08-025-038-33/C
; Sequence 33, Application US/08025038
; Patent No. 5545526
; GENERAL INFORMATION:
; APPLICANT: BAXTER-LOWE, Lee-Ann
; TITLE OF INVENTION: Method For HLA Typing
; NUMBER OF SEQUENCES: 46
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Foley & Lardner
; STREET: 777 E. Wisconsin Avenue
; CITY: Milwaukee
; STATE: Wisconsin
; COUNTRY: USA
; ZIP: 53202-5367
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/025,038
; FILING DATE: 19930301
; CLASSIFICATION: 435
```

```
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/544,218
; FILING DATE: 27-JUN-1990
; ATTORNEY/AGENT INFORMATION:
; NAME: Meyers, Philip G.
; REGISTRATION NUMBER: 30,478
; REFERENCE/DOCKET NUMBER: 204 854
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (414)289-3761
; TELEFAX: (414)289-3791
; INFORMATION FOR SEQ ID NO: 33:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 269 base pairs
; TYPE: NUCLEIC ACID
; STRANDEDNESS: single
; TOPOLOGY: linear
;
US-08-025-038-33

Query Match      100.0%; Score 11; DB 1; Length 269;
Best Local Similarity 100.0%; Pred. No. 3.2e+02;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 tctcaccacc 11
        |||
Db      249 TCTCACCAC 239
```

Search completed: July 29, 2002, 23:56:14  
Job time: 4/55 sec



11

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 30, 2002, 00:01:18 : Search time 285.14 Seconds  
(without alignments)  
36.128 Million cell updates/sec

Title: US-09-530-663B-16

Perfect score: 6

Sequence: 1 cgcgc 6

Scoring table:

Gapop 10.0, Gapext 1.0

Searched: 1736436 seqs, 858457221 residues

Total number of hits satisfying chosen parameters: 3472872

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

N\_Geneseq\_032802:\*

1: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1980.DAT:\*

2: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1981.DAT:\*

3: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1982.DAT:\*

4: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1983.DAT:\*

5: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1984.DAT:\*

6: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1985.DAT:\*

7: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1986.DAT:\*

8: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1987.DAT:\*

9: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1988.DAT:\*

10: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1989.DAT:\*

11: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1990.DAT:\*

12: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1991.DAT:\*

13: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1992.DAT:\*

14: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1993.DAT:\*

15: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1994.DAT:\*

16: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1995.DAT:\*

17: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT:\*

18: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT:\*

19: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT:\*

20: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT:\*

21: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT:\*

22: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT:\*

23: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2001B.DAT:\*

24: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result NO.	Score	Query Match Length	ID	Description
1	6	100.0	10	AAQ99824
2	6	100.0	10	AAQ99824
3	6	100.0	10	AAQ99824
4	6	100.0	10	AAQ99824
5	6	100.0	10	AAQ99824
6	6	100.0	10	AAQ99824
7	6	100.0	10	AAQ99824
8	6	100.0	10	AAQ99824
9	6	100.0	10	AAQ99824

10	6	100.0	10	21	AAZ79193
11	6	100.0	10	21	AAZ79337
12	6	100.0	10	21	AAZ81469
13	6	100.0	10	21	AAZ82818
14	6	100.0	10	21	AAZ83972
15	6	100.0	10	21	AAZ84187
16	6	100.0	10	21	AAZ85105
17	6	100.0	10	21	AAZ86191
18	6	100.0	10	22	AAH63480
19	6	100.0	10	22	AAH63759
20	6	100.0	10	22	AAQ91836
21	6	100.0	10	22	AAQ93023
22	6	100.0	10	22	AAQ93902
23	6	100.0	10	22	AAQ934378
24	6	100.0	10	22	AAQ93080
25	6	100.0	10	22	AAQ93094
26	6	100.0	10	22	AAQ93249
27	6	100.0	10	22	AAQ93594
28	6	100.0	10	22	AAQ93695
29	6	100.0	10	22	AAQ93655
30	6	100.0	10	22	AAQ93153
31	6	100.0	10	22	AAQ939516
32	6	100.0	10	22	AAQ93708
33	6	100.0	10	22	AAQ94052
34	6	100.0	10	22	AAQ94061
35	6	100.0	10	22	AAQ941006
36	6	100.0	10	22	AAQ941763
37	6	100.0	10	22	AAQ941764
38	6	100.0	10	22	AAQ941765
39	6	100.0	10	22	AAQ941766
40	6	100.0	10	22	AAQ942678
41	6	100.0	10	22	AAQ942824
42	6	100.0	10	22	AAQ942812
43	6	100.0	12	13	AAQ94014
44	6	100.0	12	15	AAQ95647
45	6	100.0	12	17	AAQ95204

#### ALIGNMENTS

RESULT 1	
AAQ99824/c	
ID	AAQ99824 standard; cDNA; 10 BP.
XX	
XX	AAQ99824;
AC	
XX	
DF	06-MAR-1996 (first entry)
XX	
DE	Lobliolly pine fusiform rust disease resistance marker OPC6 primer.
XX	
KW	Lobliolly pine; Pinus taeda; fusiform rust disease; resistance marker;
KW	Cronartium quercuum f.sp. fusiforme; Cqf; RAPD genetic marker;
KW	Random amplified polymorphic DNA analysis; woody perennial plant;
KW	family selection; pedigree; mapping; primer; ss.
OS	Synthetic.
XX	
PN	WO9519697-A1.
XX	
PD	27-JUL-1995.
XX	
PF	19-JAN-1995; 95MO-US00677.
XX	
PR	21-JAN-1994; 94US-0184567.
XX	
PA	(UYNC-) UNIV NORTH CAROLINA STATE.
XX	
PI	Grattapaglia D, O'Malley DM, Sederoff RR;
XX	WPI, 1995-269212/35.
DR	
XX	Determn. of heritable oligogenic traits in woody plants by genomic

PT mapping of multiple markers in a two generation plant family - used  
 PT to select plants with desired characteristics for breeding.

XX Example 5; Page 31; 103pp; English.

CC RAPD analysis was used to study resistance to particular strains of  
 CC Cronartium quercuum f.sp. fusiforme (Cqf), the causative agent of  
 CC fusiform rust disease, in loblolly pine (Pinus taeda). A putative  
 CC heterozygous mother tree (clone 10-5) and two open pollinated  
 CC daughters (half-sib clones 152-231 and 152-257) were crossed to a  
 CC highly susceptible pollen parent. Progeny were challenged with  
 CC inoculum from various aeciospore lines. It was found that the  
 CC marker amplified by the 10-mer primer in AAQ9824 was predictive of  
 CC resistance to inoculation with single Aeciospore line 2-36 in  
 CC clone 152-231 progeny. These and other results showed that resistance  
 CC to fusiform rust disease in loblolly pine is under oligogenic  
 CC control which can be mapped using genetic markers, using only a  
 CC two-generation pedigree.

XX Sequence 10 BP; 3 A; 3 C; 3 G; 1 T; 0 other;

Query Match 100.0%; Score 6; DB 16; Length 10;  
 Best Local Similarity 100.0%; Pred.No. 1.1e+05;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ccgcttc 6  
 |||||  
 Db 6 CCCTTC 1

RESULT 2  
 AA59797/C  
 ID AA59797 standard; DNA; 10 BP.

XX AA59797;

DT 28-JUL-1999 (first entry)

DE Primer OPC6 for fusiform rust disease resistance marker.

XX Genetic marker; genetic locus; resistance; fusiform rust disease;

KW tree family; Pinus; PCR primer; ss.

XX Synthetic.

OS US5908978-A.

PN 01-JUN-1999.

XX 18-OCT-1995; 95US-0545253.

PR 18-OCT-1995; 95US-0545253.

PR 21-JAN-1994; 94US-0184567.

XX (UNNC-) UNIV NORTH CAROLINA STATE.

PI Amerson HV, Grattapaglia D, Kuhlman EG, O'Malley DM;

PI Sederoff RR, Wilcox P;

XX WPI: 1999-347038/29.

DR Identifying resistance to fusiform rust disease in trees of the

PT genus Pinus

XX Example 5; Column 23; 69pp; English.

CC The specification describes a method of identifying a genetic marker  
 CC associated with a genetic locus conferring resistance to fusiform  
 CC rust disease in a family of trees of the genus Pinus. The method  
 CC comprises obtaining a sexually mature Pinus parent tree exhibiting  
 CC resistance to fusiform rust disease, obtaining a plurality of progeny  
 CC trees of the parent by self or cross-pollinations, assessing multiple

CC progeny trees for a number of genetic markers, identifying genetic  
 CC markers segregating in a Mendelian ratio and showing linkage with other  
 CC genetic markers, measuring resistance to fusiform rust disease in  
 CC multiple progeny trees and correlating the presence of resistance to  
 CC fusiform rust disease with at least one marker identified in the  
 CC previous step. The method is useful for determining the genetic basis  
 CC of resistance to fusiform rust disease and for producing trees of the  
 CC Pinus genus that are resistant to the disease. The present primer was  
 CC used in the method of the invention to identify and amplify resistance  
 CC markers.

XX Sequence 10 BP; 3 A; 3 C; 3 G; 1 T; 0 other;

Query Match 100.0%; Score 6; DB 20; Length 10;  
 Best Local Similarity 100.0%; Pred.No. 1.1e+05;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ccgcttc 6  
 |||||  
 Db 6 CCCTTC 1

RESULT 3  
 AA23376  
 ID AA23376 standard; DNA; 10 BP.

XX AA23376;

DT 17-JUN-1999 (first entry)

DE HLA-A, HLA-B, HLA-C polymorphism-specific primer E3b1 DNA.

XX Primer; polymorphism; HLA-A; HLA-B; HLA-C; detection; probe array;

KW hybridisation pattern; ss.

XX Synthetic.

PN NL1006733-C2.

XX 09-FEB-1999.

PF 07-AUG-1997; 97NL-1006733.

PR 07-AUG-1997; 97NL-1006733.

XX (LEES/) LEE S H.

PI Lee SH;

XX WPI: 1999-213125/18.

PT Hybridisation assay for identifying alleles - using array of  
 PT polymorphism-specific oligo:nucleotide probes

XX Claim 16; Page 26; 67pp; Dutch.

CC This invention describes a method for identifying an HLA-A, HLA-B or  
 CC HLA-C allele by detecting polymorphisms using the primers  
 CC AA23371-X23380. The method involves (a) preparing single-stranded  
 CC nucleic acid molecules corresponding in sequence to the portion of the  
 CC allele containing the target polymorphisms, (b) labelling the  
 CC single-stranded nucleic acid molecules, immobilising oligonucleotide  
 CC probes, each specific for a known polymorphism, on a support and (c)  
 CC detecting fully complementary duplexes formed between the labelled  
 CC single-stranded nucleic acid molecules and the immobilised probes. The  
 CC method is used for determining differences and correspondences in  
 CC polymorphisms between individuals, tissues or organs by comparing  
 CC hybridisation patterns produced by the above method. The method uses  
 CC probe arrays rather than requiring sequential hybridisation and removal  
 CC of individual labelled probes.

XX Sequence 10 BP; 0 A; 5 C; 1 G; 4 T; 0 other;



Query Match 100.0%; Score 6; DB 20; Length 10;  
 Best Local Similarity 100.0%; Pred. No. 1.1e+05;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 ccgttc 6  
 |||||  
 DB 5 ccgttc 10

## RESULT 4

AAC81828  
 ID AAC81828 standard; DNA; 10 BP.

AC AAC81828;

DT 22-FEB-2001 (first entry)

DE Gerbera flavone synthase FNSII primer decamer 6.

XX Gerbera: transgenic plant; flavone synthase II; FNSII; anticancer;

KW immunomodulator; naringenin; apigenin; ornamental plant; flower colour;  
 KM pharmaceutical; cancer; treatment; primer; ss.

OS Gerbera hybrida.

PN DEL918365-A1.

PD 26-OCT-2000.

PF 22-APR-1999; 99DE-1018365.

PR 22-APR-1999; 99DE-1018365.

PA (MART/) MARTENS S.  
 (FORK/) FORKMANN G.

PI Martens S, Forkmann G;

DR WPI: 2000-648348/63.

XX New nucleic acid encoding flavone synthase II, useful e.g. for  
 PT producing transgenic plants with altered flower color or flavone  
 PT content -

PS Example 4; Page 22; 40pp; German.

CC This invention describes a novel nucleic acid (I) that encodes flavone  
 CC synthase II (FNSII) which has anticancer and immunomodulatory activity.  
 CC FNSII catalyses conversion of naringenin to apigenin. (I) is used to

CC produce transgenic ornamental plants that have targeted alterations in  
 CC flower color, also altered content/distribution of flavones in leaves,  
 CC flowers and other tissues, e.g. increased resistance properties or  
 CC symbiotic capacity. FNSII expressed by (I) is used in synthesis of  
 CC flavones that are useful as pharmaceuticals, e.g. in cancer treatment,  
 CC as biologically active substances, e.g. to improve the immune defence  
 CC system. Oligonucleotide fragments of (I) are used as probes and primers,  
 CC or as antisense or ribozyme agents for regulating expression of (II).

XX Sequence 10 BP; 0 A; 4 C; 4 G; 2 T; 0 other;

Query Match 100.0%; Score 6; DB 21; Length 10;

Best Local Similarity 100.0%; Pred. No. 1.1e+05;

Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 ccgttc 6  
 |||||  
 DB 3 ccgttc 8

## RESULT 5

AAC81843  
 ID AAC81843 standard; DNA; 10 BP.

AC AAC81843;

DT 22-FEB-2001 (first entry)

DE Gerbera flavone synthase FNSII haem-binding DNA fragment D6.

KW Gerbera: transgenic plant; flavone synthase II; FNSII; anticancer;  
 KM immunomodulator; naringenin; apigenin; ornamental plant; flower colour;  
 KM pharmaceutical; cancer; treatment; ss.

OS Gerbera hybrida.

PN DEL918365-A1.

PD 26-OCT-2000.

PF 22-APR-1999; 99DE-1018365.

PR 22-APR-1999; 99DE-1018365.

PA (MART/) MARTENS S.  
 (FORK/) FORKMANN G.

PI Martens S, Forkmann G;

DR WPI: 2000-648348/63.

XX New nucleic acid encoding flavone synthase II, useful e.g. for  
 PT producing transgenic plants with altered flower color or flavone  
 PT content -

PS Disclosure; Fig 4B; 40pp; German.

CC This invention describes a novel nucleic acid (I) that encodes flavone  
 CC synthase II (FNSII) which has anticancer and immunomodulatory activity.  
 CC FNSII catalyses conversion of naringenin to apigenin. (I) is used to

CC produce transgenic ornamental plants that have targeted alterations in  
 CC flower color, also altered content/distribution of flavones in leaves,  
 CC flowers and other tissues, e.g. increased resistance properties or  
 CC symbiotic capacity. FNSII expressed by (I) is used in synthesis of  
 CC flavones that are useful as pharmaceuticals, e.g. in cancer treatment,  
 CC as biologically active substances, e.g. to improve the immune defence  
 CC system. Oligonucleotide fragments of (I) are used as probes and primers,  
 CC or as antisense or ribozyme agents for regulating expression of (II).

XX Sequence 10 BP; 0 A; 4 C; 4 G; 2 T; 0 other;

Query Match 100.0%; Score 6; DB 21; Length 10;

Best Local Similarity 100.0%; Pred. No. 1.1e+05;

Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 ccgttc 6  
 |||||  
 DB 3 ccgttc 8

## RESULT 6

AAZ50866  
 ID AAZ50866 standard; DNA; 10 BP.

AC AAZ50866;

DT 31-MAY-2000 (first entry)

DE Primer AP11 to identify tobacco salicylic acid inducible genes.

XX Tobacco plant; salicylic acid inducible gene; fungal pathogen;  
 KM SA-inducible gene; transgenic plant; pathogen resistance; PCR primer; ss.  
 XX

OS Nicotiana tabacum.  
 XX WO200008186-A1.  
 XX 17-FEB-2000.  
 PD  
 XX 02-AUG-1999; 99WO-EP05581.  
 PF  
 XX 03-AUG-1998; 98US-0095187.  
 PR  
 XX (MOGE-) MOGEN INT NV.  
 PA  
 XX Stuiver MH, Jepson I, Horvath DM, Chua N;  
 PI WPI; 2000-205725/18.  
 DR  
 XX Novel salicylic acid inducible genes from tobacco plants, useful for  
 PT making transgenic plants with enhanced pathogenic resistance -  
 XX Example 1; Page 53; 57pp; English.  
 PS  
 XX The patent discloses fifteen new salicylic acid (SA) inducible genes from  
 CC Nicotiana tabacum, nine of which were subcloned and sequenced, based on  
 CC different kinetics of induction response, these genes were classified  
 CC into four categories, class I, II, III and IV response genes. The  
 CC SA-inducible genes are useful for making transgenic plants with enhanced  
 CC pathogen resistance. The plants incorporating these genes show reduced  
 CC susceptibility to fungal pathogens. The present sequence is an  
 CC upstream primer APL1 used in differential display PCR reactions along  
 CC with downstream primers T12MG or T12MC to identify tobacco  
 CC SA-inducible genes.  
 XX  
 SQ Sequence 10 BP; 2 A; 4 C; 2 G; 2 T; 0 other;

Query Match 100.0%; Score 6; DB 21; Length 10;  
 Best Local Similarity 100.0%; Pred No. 1.1e+05;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1 ccgttc 6  
 |||||  
 Db 5 ccgttc 10

RESULT 7  
 AA277645/c  
 ID AA277645 standard; DNA; 10 BP.  
 XX AA277645;  
 AC  
 XX 10-APR-2000 (first entry)  
 DT  
 XX Human dendritic cell SAGE tag, SEQ ID NO:73.  
 DE  
 XX SAGE tag; serial analysis of gene expression; antigen-presenting cell;  
 KW APC; monocyte-derived dendritic cell; differential gene expression;  
 KW Immunostimulatory cofactor; costimulatory factor; CTL;  
 KW cytotoxic T-lymphocyte; tumour antigen; immunotherapy; anticancer; ss.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO965924-A2.  
 PD  
 XX 23-DEC-1999.  
 PD  
 XX 18-JUN-1999; 99WO-US13800.  
 PF  
 XX 19-JUN-1998; 98US-0089833.  
 PR 19-JUN-1998; 98US-0089844.  
 PR 19-JUN-1998; 98US-0089853.  
 PR 19-JUN-1998; 98US-0089878.  
 PR 19-JUN-1998; 98US-0089891.  
 PR 19-JUN-1998; 98US-0089992.

PR 19-JUN-1998; 98US-0089993.  
 PR 19-JUN-1998; 98US-0089994.  
 PR 19-JUN-1998; 98US-0089997.  
 PR 19-JUN-1998; 98US-0089999.  
 PR 19-JUN-1998; 98US-0090000.  
 PR 19-JUN-1998; 98US-0090003.  
 PR 19-JUN-1998; 98US-0090036.  
 PR 19-JUN-1998; 98US-0090039.  
 PR 19-JUN-1998; 98US-0090040.  
 PR 19-JUN-1998; 98US-0090041.  
 PR 19-JUN-1998; 98US-0090042.  
 PR 19-JUN-1998; 98US-0090043.  
 PR 19-JUN-1998; 98US-0090044.  
 PR 19-JUN-1998; 98US-0090045.  
 PR 19-JUN-1998; 98US-0090047.  
 PR 19-JUN-1998; 98US-0090048.  
 PR 19-JUN-1998; 98US-0090072.  
 PR 19-JUN-1998; 98US-0090076.  
 PR 19-JUN-1998; 98US-0090077.  
 PR 19-JUN-1998; 98US-0090078.  
 PR 19-JUN-1998; 98US-0090079.  
 PR 19-JUN-1998; 98US-0090080.  
 PR 08-DEC-1998; 98US-011715.  
 XX  
 PA (GEN2 ) GENZYME CORP.  
 PA (ROBE/) ROBERTS B L.  
 PA (SHAN/) SHANKARA S.  
 PI  
 XX Roberts BL, Shankara S;  
 XX WPI; 2000-106077/09.  
 DR  
 XX Isolated polynucleotides differentially expressed in antigen-presenting  
 PT cells, useful in gene vaccines against cancer -  
 XX Claim 1; Page 65; 130pp; English.

Sequences AA277573-279709 represent SAGE (serial analysis of gene  
 expression) tags used to identify mRNA transcripts encoding  
 CC immunostimulatory cofactor proteins which are preferentially or  
 CC differentially expressed in monocyte-derived dendritic cells compared  
 CC with monocytes. Some of the transcripts correspond to known genes or  
 CC ESTs (expressed sequence tags) which were previously unknown to be  
 CC preferentially or differentially expressed in dendritic cells, while  
 CC other transcripts correspond to novel genes. Antigen-presenting cell  
 CC (APC)-associated costimulatory factors play an important role in the  
 CC activation of the cytotoxic immune response, particularly against tumour  
 CC cells. Tumour antigen presentation via the MHC (major histocompatibility  
 CC complex) and subsequent recognition by T-cell receptors is alone  
 CC insufficient to activate a robust cytotoxic immune response that can  
 CC lyse the tumour cells; immunostimulatory cofactors also being required  
 CC for efficient activation of cytotoxic T-lymphocytes (CTLs). Nucleic acid  
 CC sequences identified using the SAGE tags have several potential uses.  
 CC They may be used in vaccines to induce an immune response, particularly  
 CC against a tumour antigen; to modulate the genotype of an APC; to screen  
 CC for agents that modulate expression of differentially expressed genes in  
 CC an APC; and as hybridisation probes/amplification primers for the  
 CC diagnosis, prognosis and monitoring of diseases related to abnormal  
 CC expression of these genes. Detection of the dendritic cell  
 CC differentially expressed genes, or of their encoded proteins, can be used  
 CC to identify cells as belonging to the monocyte lineage. Cells containing  
 CC these genes can be used in active immunotherapy (or to stimulate  
 CC production of a population of antigen-specific effector cells) and  
 CC vectors containing them are used in gene therapy. Co-administration of  
 CC tumour antigens and APC-associated costimulatory factors ensures adequate  
 CC antigen presentation to endogenous APCs and upregulates the APCs for the  
 CC secretion of co-stimulatory signals, migration to T cell-rich sites,  
 CC recruitment of immune effector cells.  
 CC  
 SQ Sequence 10 BP; 3 A; 1 C; 6 G; 0 U; 0 other;

Query Match 100.0%; Score 6; DB 21; Length 10;  
Best Local Similarity 100.0%; Pred. No. 1.1e+05;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ccgttc 6  
      |||||  
Db 8 CCGTTC 3

RESULT 8  
AAZ77674  
ID AAZ77674 standard; DNA; 10 BP.  
XX  
AC AAZ77674;  
XX  
DT 10-APR-2000 (first entry)  
XX  
DE Human dendritic cell SAGE tag, SEQ ID NO:102.  
XX  
KW SAGE tag; serial analysis of gene expression; antigen-presenting cell;  
KW APC; monocyte-derived dendritic cell; differential gene expression;  
KW immunostimulatory cofactor; costimulatory factor; CTL;  
KW cytotoxic T-lymphocyte; tumour antigen; immunotherapy; anticancer; ss.  
XX  
OS Homo sapiens.  
XX  
PN WO9965924-A2.  
XX  
PD 23-DEC-1999.  
XX  
PF 18-JUN-1999; 99WO-US13800.  
XX  
PR 19-JUN-1998; 98US-0089833.  
PR 19-JUN-1998; 98US-0089844.  
PR 19-JUN-1998; 98US-0089853.  
PR 19-JUN-1998; 98US-0089878.  
PR 19-JUN-1998; 98US-0089991.  
PR 19-JUN-1998; 98US-0089992.  
PR 19-JUN-1998; 98US-0089993.  
PR 19-JUN-1998; 98US-0089994.  
PR 19-JUN-1998; 98US-0089997.  
PR 19-JUN-1998; 98US-0089999.  
PR 19-JUN-1998; 98US-0090000.  
PR 19-JUN-1998; 98US-0090035.  
PR 19-JUN-1998; 98US-0090036.  
PR 19-JUN-1998; 98US-0090039.  
PR 19-JUN-1998; 98US-0090040.  
PR 19-JUN-1998; 98US-0090041.  
PR 19-JUN-1998; 98US-0090042.  
PR 19-JUN-1998; 98US-0090043.  
PR 19-JUN-1998; 98US-0090044.  
PR 19-JUN-1998; 98US-0090045.  
PR 19-JUN-1998; 98US-0090047.  
PR 19-JUN-1998; 98US-0090048.  
PR 19-JUN-1998; 98US-0090072.  
PR 19-JUN-1998; 98US-0090076.  
PR 19-JUN-1998; 98US-0090077.  
PR 19-JUN-1998; 98US-0090078.  
PR 19-JUN-1998; 98US-0090079.  
PR 19-JUN-1998; 98US-0090080.  
PR 08-DEC-1998; 98US-0111715.  
XX  
PA (GENZ ) GENZYME CORP.  
PA (ROBE/) ROBERTS B L.  
PA (SHAN/) SHANKARA S.  
XX  
PI Roberts BL, Shankara S;  
XX  
DR WPI: 2000-106077/09.  
XX  
PT Isolated polynucleotides differentially expressed in antigen-presenting  
PT cells, useful in gene vaccines against cancer -  
XX

PS Claim 1; Page 66; 130pp; English.  
XX  
CC Sequences AAZ7573-779709 represent SAGE (serial analysis of gene  
CC expression) tags used to identify mRNA transcripts encoding  
CC immunostimulatory cofactor proteins which are preferentially or  
CC differentially expressed in monocyte-derived dendritic cells compared  
CC with monocytes. Some of the transcripts correspond to known genes or  
CC ESTs (expressed sequence tags) which were previously unknown to be  
CC preferentially or differentially expressed in dendritic cells, while  
CC other transcripts correspond to novel genes. Antigen-presenting cell  
CC (APC)-associated costimulatory factors play an important role in the  
CC activation of the cytotoxic immune response, particularly against tumour  
CC cells. Tumour antigen presentation via the MHC (major histocompatibility  
CC complex) and subsequent recognition by T-cell receptors is alone  
CC insufficient to activate a robust cytotoxic immune response that can  
CC lyse the tumour cells, immunostimulatory cofactors also being required  
CC for efficient activation of cytotoxic T-lymphocytes (CTLs). Nucleic acid  
CC sequences identified using the SAGE tags have several potential uses.  
CC They may be used in vaccines to induce an immune response, particularly  
CC against a tumour antigen; to modulate the genotype of an APC; to screen  
CC for agents that modulate expression of differentially expressed genes in  
CC an APC; and as hybridisation probes/amplification primers for the  
CC diagnosis, prognosis and monitoring of diseases related to abnormal  
CC expression of these genes. Detection of the dendritic cell  
CC differentially expressed genes, or of their encoded proteins, can be used  
CC to identify cells as belonging to the monocyte lineage. Cells containing  
CC these genes can be used in active immunotherapy (or to stimulate  
CC production of a population of antigen-specific effector cells) and  
CC vectors containing them are used in gene therapy. Co-administration of  
CC tumour antigens and APC-associated costimulatory factors ensures adequate  
CC antigen presentation to endogenous APCs and upregulates the APCs for the  
CC presentation of co-stimulatory signals, migration to T cell-rich sites,  
CC secretion of T cell growth factors and secretion of chemokines for  
CC recruitment of immune effector cells.  
XX  
SQ Sequence 10 BP; 1 A; 3 C; 2 G; 4 T; 0 other;

Qy 1 ccgttc 6  
      |||||  
Db 2 ccgttc 7

RESULT 9  
AAZ78756/C  
ID AAZ78756 standard; DNA; 10 BP.  
XX  
AC AAZ78756;  
XX  
DT 10-APR-2000 (first entry)  
XX  
DE Human dendritic cell SAGE tag, SEQ ID NO:1184.  
XX  
DE  
XX  
KW SAGE tag; serial analysis of gene expression; antigen-presenting cell;  
KW APC; monocyte-derived dendritic cell; differential gene expression;  
KW immunostimulatory cofactor; costimulatory factor; CTL;  
KW cytotoxic T-lymphocyte; tumour antigen; immunotherapy; anticancer; ss.  
XX  
XX  
OS Homo sapiens.  
XX  
PN WO9965924-A2.  
XX  
PD 23-DEC-1999.  
XX  
PF 18-JUN-1999; 99WO-US13800.  
XX  
PR 19-JUN-1998; 98US-0089833.  
PR 19-JUN-1998; 98US-0089844.  
PR 19-JUN-1998; 98US-0089853.

PR 19-JUN-1998; 980S-0089878.  
PR 19-JUN-1998; 980S-0089991.  
PR 19-JUN-1998; 980S-0089992.  
PR 19-JUN-1998; 980S-0089993.  
PR 19-JUN-1998; 980S-0089994.  
PR 19-JUN-1998; 980S-0089997.  
PR 19-JUN-1998; 980S-0089999.  
PR 19-JUN-1998; 980S-0090000.  
PR 19-JUN-1998; 980S-0090035.  
PR 19-JUN-1998; 980S-0090036.  
PR 19-JUN-1998; 980S-0090039.  
PR 19-JUN-1998; 980S-0090040.  
PR 19-JUN-1998; 980S-0090041.  
PR 19-JUN-1998; 980S-0090042.  
PR 19-JUN-1998; 980S-0090043.  
PR 19-JUN-1998; 980S-0090044.  
PR 19-JUN-1998; 980S-0090045.  
PR 19-JUN-1998; 980S-0090047.  
PR 19-JUN-1998; 980S-0090048.  
PR 19-JUN-1998; 980S-0090072.  
PR 19-JUN-1998; 980S-0090076.  
PR 19-JUN-1998; 980S-0090077.  
PR 19-JUN-1998; 980S-0090078.  
PR 19-JUN-1998; 980S-0090079.  
PR 19-JUN-1998; 980S-0090080.  
PR 08-DEC-1998; 980S-0111715.  
XX  
PA (GENZ ) GENZYME CORP.  
PA (ROBE/) ROBERTS B L.  
PA (SHAN/) SHANKARA S.  
XX  
PI Roberts BL, Shankara S;  
XX  
DR WPI: 2000-106077/09.  
XX  
PT Isolated polynucleotides differentially expressed in antigen-presenting  
XX cells, useful in gene vaccines against cancer .  
XX  
PS Claim 1: Page 98; 130pp; English.  
XX  
CC Sequences AA27573-279709 represent SAGE (serial analysis of gene  
CC expression) tags used to identify mRNA transcripts encoding  
CC immunostimulatory cofactor proteins which are preferentially or  
CC differentially expressed in monocyte-derived dendritic cells compared  
CC with monocytes. Some of the transcripts correspond to known genes or  
CC ESTs (expressed sequence tags) which were previously unknown to be  
CC preferentially or differentially expressed in dendritic cells, while  
CC other transcripts correspond to novel genes. Antigen-presenting cell  
CC (APC)-associated costimulatory factors play an important role in the  
CC activation of the cytotoxic immune response, particularly against tumour  
CC cells. Tumour antigen presentation via the MHC (major histocompatibility  
CC complex) and subsequent recognition by T-cell receptors is alone  
CC insufficient to activate a robust cytotoxic immune response that can  
CC lyse the tumour cells, immunostimulatory cofactors also being required  
CC for efficient activation of cytotoxic T-lymphocytes (CTLs). Nucleic acid  
CC sequences identified using the SAGE tags have several potential uses.  
CC They may be used in vaccines to induce an immune response, particularly  
CC against a tumour antigen; to modulate the genotype of an APC; to screen  
CC for agents that modulate expression of differentially expressed genes in  
CC an APC; and as hybridisation probes/amplification primers for the  
CC diagnosis, prognosis and monitoring of diseases related to abnormal  
CC expression of these genes. Detection of the dendritic cell  
CC differentially expressed genes, or of their encoded proteins, can be used  
CC to identify cells as belonging to the monocyte lineage. Cells containing  
CC these genes can be used in active immunotherapy (or to stimulate  
CC production of a population of antigen-specific effector cells) and  
CC vectors containing them are used in gene therapy. Co-administration of  
CC tumour antigens and APC-associated costimulatory factors ensures adequate  
CC antigen presentation to endogenous APCs and upregulates the APCs for the  
CC presentation of co-stimulatory signals, migration to T cell-rich sites,  
CC secretion of T cell growth factors and secretion of chemokines for  
CC recruitment of immune effector cells.  
XX

SQL Sequence 10 BP; 3 A; 2 C; 5 G; 0 U; 0 other;  
Query Match 100.0%; Score 6; DB 21; Length 10;  
Best Local Similarity 100.0%; Pred. No. 1.1e+05;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Qy 1 ccgctc 6  
Db 9 CCCTTC 4  
RESULT 10  
AA279193/C  
ID AA279193 standard; DNA; 10 BP.  
XX  
AC AA279193;  
XX  
DT 10-APR-2000 (first entry)  
XX  
DE Human dendritic cell SAGE tag, SEQ ID NO:1621.  
XX  
KW SAGE tag; serial analysis of gene expression; antigen-presenting cell;  
KW APC; monocyte-derived dendritic cell; differential gene expression;  
KW immunostimulatory cofactor; costimulatory factor; CTL;  
KW cytotoxic T-lymphocyte; tumour antigen; immunotherapy; anticancer; ss.  
XX  
OS Homo sapiens.  
XX  
PN WO965924-A2.  
XX  
PD 23-DEC-1999.  
XX  
PF 18-JUN-1999; 99WO-US13800.  
XX  
PR 19-JUN-1998; 980S-0089833.  
PR 19-JUN-1998; 980S-0089844.  
PR 19-JUN-1998; 980S-0089853.  
PR 19-JUN-1998; 980S-0089878.  
PR 19-JUN-1998; 980S-0089997.  
PR 19-JUN-1998; 980S-0089999.  
PR 19-JUN-1998; 980S-0090035.  
PR 19-JUN-1998; 980S-0090036.  
PR 19-JUN-1998; 980S-0090039.  
PR 19-JUN-1998; 980S-0090040.  
PR 19-JUN-1998; 980S-0090041.  
PR 19-JUN-1998; 980S-0090042.  
PR 19-JUN-1998; 980S-0090043.  
PR 19-JUN-1998; 980S-0090044.  
PR 19-JUN-1998; 980S-0090045.  
PR 19-JUN-1998; 980S-0090047.  
PR 19-JUN-1998; 980S-0090048.  
PR 19-JUN-1998; 980S-0090072.  
PR 19-JUN-1998; 980S-0090076.  
PR 19-JUN-1998; 980S-0090077.  
PR 19-JUN-1998; 980S-0090078.  
PR 19-JUN-1998; 980S-0090079.  
PR 19-JUN-1998; 980S-0090080.  
PR 08-DEC-1998; 980S-0111715.  
XX  
PA (GENZ ) GENZYME CORP.  
PA (ROBE/) ROBERTS B L.  
PA (SHAN/) SHANKARA S.  
XX  
PI Roberts BL, Shankara S;  
XX  
DR WPI: 2000-106077/09.  
XX



CC secretion of T cell growth factors and secretion of chemokines for  
CC recruitment of immune effector cells.  
XX  
SQ Sequence 10 BP; 1 A; 3 C; 3 G; 3 T; 0 other;

Query Match 100.0%; Score 6; DB 21; Length 10;  
Best Local Similarity 100.0%; Pred. No. 1.1e+05;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 ccgttc 6  
111111  
DB 1 ccgttc 6

## RESULT 12

AAZ81469  
ID AAZ81469 standard; DNA; 10 BP.

AC AAZ81469;

DT 07-APR-2000 (first entry)

DE Metastatic breast tumour cell upregulated transcript tag #703.

XX  
XX Human; metastatic breast tumour tissue; breast cancer; tag; primer;  
KW non-metastatic breast tumour tissue; gene therapy; anticancer;  
KW antimetastatic; vaccine; diagnosis; ss.

XX Homo sapiens.

XX WO965928-A2.

PN 23-DEC-1999.

PD 18-JUN-1999; 99WO-US13647.

PF 19-JUN-1998; 98US-0089853.

PR 19-JUN-1998; 98US-0089997.

PR 19-JUN-1998; 98US-0090039.

PR 19-JUN-1998; 98US-0090040.

PR 19-JUN-1998; 98US-0090041.

XX (GENZ ) GENZYME CORP.

PA (ROBE/) ROBERTS B L.

PA (SHAN/) SHANKARA S.

PI Roberts BL, Shankara S;

DR WPI; 2000-106079/09.

XX Isolated polynucleotides differentially expressed between metastatic

PT and non-metastatic breast cancer cells, useful for diagnosis,

PT prevention and treatment of cancer -

XX Claim 1; Page 77; 219pp; English.

XX AAZ80767 to AAZ83941 represent tags corresponding to distinct

CC transcripts that are preferentially transcribed in the metastatic breast

CC tumour tissue (i.e. are upregulated in metastatic breast tumour cells).

CC AAZ83942 to AAZ86677 represent tags corresponding to distinct transcripts

CC that are preferentially transcribed in the primary or non-metastatic

CC breast tumour tissue (i.e. are downregulated in metastatic breast tumour

CC cells). These transcripts can be used for diagnosis, prognosis,

CC monitoring and treatment of breast cancer, particularly where metastatic

CC diagnosis is by standard immunoassays or hybridisation/amplification

CC reactions. Compounds that modulate expression of the transcripts are

CC potentially useful for treatment of (metastatic) breast cancer, while

CC promoters from the transcripts are used to direct expression, in selected

CC cell types, of e.g. therapeutic genes (also ribozymes or antisense

CC sequences), particularly an antigen-encoding sequence for use in gene or

CC cell-based vaccines. Polypeptides encoded by the transcripts are also

CC useful in vaccines; for diagnosing breast cancer and for raising

CC specific antibodies (Ab). Ab are used to detect the polypeptides or as  
CC therapeutic agents. Host cells that produce the polypeptides can be used  
CC to expand and isolate populations of educated, antigen-specific immune  
CC effector cells, e.g. cytotoxic T lymphocytes, and these used for  
CC adoptive immunotherapy.

XX Sequence 10 BP; 2 A; 4 C; 2 G; 2 T; 0 other;

Query Match 100.0%; Score 6; DB 21; Length 10;  
Best Local Similarity 100.0%; Pred. No. 1.1e+05;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 ccgttc 6  
111111  
DB 1 ccgttc 6

## RESULT 13

AAZ82818/C  
ID AAZ82818 standard; DNA; 10 BP.

AC AAZ82818;

DT 07-APR-2000 (first entry)

DE Metastatic breast tumour cell upregulated transcript tag #2052.

XX  
XX Human; metastatic breast tumour tissue; breast cancer; tag; primer;  
KW non-metastatic breast tumour tissue; gene therapy; anticancer;  
KW antimetastatic; vaccine; diagnosis; ss.

XX Homo sapiens.

XX WO965928-A2.

PN 23-DEC-1999.

PD 18-JUN-1999; 99WO-US13647.

PF 19-JUN-1998; 98US-0089853.

PR 19-JUN-1998; 98US-0089997.

PR 19-JUN-1998; 98US-0090039.

PR 19-JUN-1998; 98US-0090040.

PR 19-JUN-1998; 98US-0090041.

XX (GENZ ) GENZYME CORP.

PA (ROBE/) ROBERTS B L.

PA (SHAN/) SHANKARA S.

PI Roberts BL, Shankara S;

DR WPI; 2000-106079/09.

XX Isolated polynucleotides differentially expressed between metastatic

PT and non-metastatic breast cancer cells, useful for diagnosis,

PT prevention and treatment of cancer -

XX Claim 1; Page 114; 219pp; English.

XX AAZ80767 to AAZ83941 represent tags corresponding to distinct

CC transcripts that are preferentially transcribed in the metastatic breast

CC tumour tissue (i.e. are upregulated in metastatic breast tumour cells).

CC AAZ83942 to AAZ86677 represent tags corresponding to distinct transcripts

CC that are preferentially transcribed in the primary or non-metastatic

CC breast tumour tissue (i.e. are downregulated in metastatic breast tumour

CC cells). These transcripts can be used for diagnosis, prognosis,

CC monitoring and treatment of breast cancer, particularly where metastatic

CC diagnosis is by standard immunoassays or hybridisation/amplification

CC reactions. Compounds that modulate expression of the transcripts are

CC potentially useful for treatment of (metastatic) breast cancer, while

CC promoters from the transcripts are used to direct expression, in selected

CC cell types, of e.g. therapeutic genes (also ribozymes or antisense

CC sequences), particularly an antigen-encoding sequence for use in gene or  
CC cell-based vaccines. Polypeptides encoded by the transcripts are also  
CC useful in vaccines; for diagnosing breast cancer and for raising  
CC specific antibodies (Ab). Ab are used to detect the polypeptides or as  
CC therapeutic agents. Host cells that produce the polypeptides can be used  
CC to expand and isolate populations of educated, antigen-specific immune  
CC effector cells, e.g. cytotoxic T lymphocytes, and these used for  
CC adoptive immunotherapy.

XX  
SQ Sequence 10 BP; 3 A; 1 C; 5 G; 1 T; 0 other;

Query Match 100.0%; Score 6; DB 21; Length 10;  
Best Local Similarity 100.0%; Pred. No. 1.1e+05;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1 ccgttc 6  
      |||||  
Db 7 CCGTTC 2

RESULT 14  
AA283972 standard; DNA; 10 BP.  
ID AA283972;  
AC AA283972;  
XX  
DT 07-APR-2000 (first entry)  
XX  
DE Metastatic breast tumour cell downregulated transcript tag #3206.  
XX  
KW Human; metastatic breast tumour tissue; breast cancer; tag; primer;  
KW non-metastatic breast tumour tissue; gene therapy; anticancer;  
KW antimetastatic; vaccine; diagnosis; ss.  
XX  
OS Homo sapiens.  
XX  
PN WO965928-A2.  
XX  
PD 23-DEC-1999.  
XX  
PE 18-JUN-1999; 99WO-US13647.  
XX  
PF 19-JUN-1998; 98US-0089853.  
PR 19-JUN-1998; 98US-0089997.  
PR 19-JUN-1998; 98US-0090039.  
PR 19-JUN-1998; 98US-0090040.  
PR 19-JUN-1998; 98US-0090041.  
XX  
PA (GENZ ) GENZYME CORP.  
PA (ROBE/) ROBERTS B L.  
PA (SHAN/) SHANKARA S.  
XX  
PI Roberts BL, Shankara S;  
XX  
DR WPI: 2000-106079/09.  
XX  
PT Isolated polynucleotides differentially expressed between metastatic  
PT and non-metastatic breast cancer cells, useful for diagnosis,  
PT prevention and treatment of cancer -  
XX  
PS Claim 1; Page 145; 219pp; English.  
XX  
AA280767 to AA283941 represent tags corresponding to distinct  
CC transcripts that are preferentially transcribed in the metastatic breast  
CC tumour tissue (i.e. are upregulated in metastatic breast tumour cells).  
CC AA283942 to AA286677 represent tags corresponding to distinct transcripts  
CC that are preferentially transcribed in the primary or non-metastatic  
CC breast tumour tissue (i.e. are downregulated in metastatic breast tumour  
CC cells). These transcripts can be used for diagnosis, prognosis,  
CC monitoring and treatment of breast cancer, particularly where metastatic.  
CC diagnosis is by standard immunoassays or hybridisation/amplification  
CC reactions. Compounds that modulate expression of the transcripts are

CC potentially useful for treatment of (metastatic) breast cancer, while  
CC promoters from the transcripts are used to direct expression, in selected  
CC cell types, of e.g. therapeutic genes (also ribozymes or antisense  
CC sequences), particularly an antigen-encoding sequence for use in gene or  
CC cell-based vaccines. Polypeptides encoded by the transcripts are also  
CC useful in vaccines; for diagnosing breast cancer and for raising  
CC specific antibodies (Ab). Ab are used to detect the polypeptides or as  
CC therapeutic agents. Host cells that produce the polypeptides can be used  
CC to expand and isolate populations of educated, antigen-specific immune  
CC effector cells, e.g. cytotoxic T lymphocytes, and these used for  
CC adoptive immunotherapy.

XX  
SQ Sequence 10 BP; 1 A; 3 C; 2 G; 4 T; 0 other;

Query Match 100.0%; Score 6; DB 21; Length 10;  
Best Local Similarity 100.0%; Pred. No. 1.1e+05;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1 ccgttc 6  
      |||||  
Db 2 ccgttc 7

RESULT 15  
AA284187  
ID AA284187 standard; DNA; 10 BP.  
XX  
AC AA284187;  
XX  
DT 07-APR-2000 (first entry)  
XX  
DE Metastatic breast tumour cell downregulated transcript tag #3421.  
XX  
KW Human; metastatic breast tumour tissue; breast cancer; tag; primer;  
KW non-metastatic breast tumour tissue; gene therapy; anticancer;  
KW antimetastatic; vaccine; diagnosis; ss.  
XX  
OS Homo sapiens.  
XX  
PN WO965928-A2.  
XX  
PD 23-DEC-1999.  
XX  
PE 18-JUN-1999; 99WO-US13647.  
XX  
PF 19-JUN-1998; 98US-0089853.  
PR 19-JUN-1998; 98US-0089997.  
PR 19-JUN-1998; 98US-0090039.  
PR 19-JUN-1998; 98US-0090040.  
PR 19-JUN-1998; 98US-0090041.  
XX  
PA (GENZ ) GENZYME CORP.  
PA (ROBE/) ROBERTS B L.  
PA (SHAN/) SHANKARA S.  
XX  
PI Roberts BL, Shankara S;  
XX  
DR WPI: 2000-106079/09.  
XX  
PT Isolated polynucleotides differentially expressed between metastatic  
PT and non-metastatic breast cancer cells, useful for diagnosis,  
PT prevention and treatment of cancer -  
XX  
PS Claim 1; Page 150; 219pp; English.  
XX  
AA280767 to AA283941 represent tags corresponding to distinct  
CC transcripts that are preferentially transcribed in the metastatic breast  
CC tumour tissue (i.e. are upregulated in metastatic breast tumour cells).  
CC AA283942 to AA286677 represent tags corresponding to distinct transcripts  
CC that are preferentially transcribed in the primary or non-metastatic  
CC breast tumour tissue (i.e. are downregulated in metastatic breast tumour  
CC cells). These transcripts can be used for diagnosis, prognosis,

CC monitoring and treatment of breast cancer, particularly where metastatic.  
CC Diagnosis is by standard immunoassays or hybridisation/amplification  
CC reactions. Compounds that modulate expression of the transcripts are  
CC potentially useful for treatment of (metastatic) breast cancer, while  
CC promoters from the transcripts are used to direct expression, in selected  
CC cell types, of e.g. therapeutic genes (also ribozymes or antisense  
CC sequences), particularly an antigen-encoding sequence for use in gene or  
CC cell-based vaccines. Polypeptides encoded by the transcripts are also  
CC useful in vaccines: for diagnosing breast cancer and for raising  
CC specific antibodies (Ab). Ab are used to detect the polypeptides or as  
CC therapeutic agents. Host cells that produce the polypeptides can be used  
CC to expand and isolate populations of educated, antigen-specific immune  
CC effector cells; e.g. cytotoxic T lymphocytes, and these used for  
CC adoptive immunotherapy.  
XX

SQ Sequence 10 BP; 0 A; 5 C; 3 G; 2 T; 0 other;

Query Match

Best Local Similarity 100.0%; Score 6; DB 21; Length 10;

Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ccgttc 6  
|||||  
Db 2 ccgttc 7

Search completed: July 30, 2002, 00:01:19  
Job time: 4930 sec



GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 29, 2002, 23:22:40 ; Search time 2542.47 Seconds  
(without alignments)  
31.852 Million cell updates/sec

Title: US-09-530-663b-16

Perfect score: 6  
Sequence: 1 ccgttc 6

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 13736207 seqs, 674847542 residues

Total number of hits satisfying chosen parameters: 27472414

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%  
Listing first 45 summaries

Database :

EST:\*  
1: em\_estbda:\*  
2: em\_esthum:\*  
3: em\_estln:\*  
4: em\_estnu:\*  
5: em\_estov:\*  
6: em\_estpl:\*  
7: em\_estro:\*  
8: em\_hlc:\*  
9: gb\_est1:\*  
10: gb\_est2:\*  
11: gb\_hlc:\*  
12: gb\_gss:\*  
13: em\_gss\_hum:\*  
14: em\_gss\_inv:\*  
15: em\_gss\_pln:\*  
16: em\_gss\_vrt:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	6	100.0	19	9	AA911671	AA911671.0149f08.s
2	6	100.0	21	12	AZ347845	AZ347845.1M0084L06
3	6	100.0	22	12	AZ811866	AZ811866.2M0078M08
4	6	100.0	23	12	AZ653863	AZ653863.1M00527D14
5	6	100.0	24	12	AZ316663	AZ316663.1M0034G22
6	6	100.0	24	12	AZ436588	AZ436588.1M0224H20
7	6	100.0	24	12	AZ446206	AZ446206.1M0242I06
8	6	100.0	24	12	AZ662500	AZ662500.1M0541G07
9	6	100.0	24	12	AZ789336	AZ789336.2M0038L17
10	6	100.0	24	12	TA143H10Q	AL467041.T. bruce1
11	6	100.0	25	9	A1199669	A1199669.q160a02.x
12	6	100.0	25	12	AZ442576	AZ442576.1M0236K13
13	6	100.0	25	12	AZ621312	AZ621312.1M0454P19
14	6	100.0	26	12	AZ372925	AZ372925.1M0125D04
15	6	100.0	26	12	AZ819947	AZ819947.2M0091E20
16	6	100.0	26	12	TA347D12P	AL493847.T. bruce1
17	6	100.0	27	10	BG927944	BG927944.HNC45-1-F

C	18	6	100.0	27	12	AZ488404	AZ488404.1M0318I11
C	19	6	100.0	27	12	AZ494628	AZ494628.1M0330E06
C	20	6	100.0	27	12	AZ809974	AZ809974.2M0074C18
C	21	6	100.0	27	12	TA6R09Q	AL451746.T. bruce1
C	22	6	100.0	28	9	A1748505	A1748505.sbs3h08.y
C	23	6	100.0	28	12	AZ780363	AZ780363.2M0017I11
C	24	6	100.0	28	12	AZ830168	AZ830168.2M0109E13
C	25	6	100.0	28	12	TA116E03P	AL465226.T. bruce1
C	26	6	100.0	29	12	AZ595520	AZ595520.1M0408M09
C	27	6	100.0	29	12	TA119F03Q	AL463260.T. bruce1
C	28	6	100.0	30	10	BG719541	BG719541.602690091
C	29	6	100.0	30	10	BG719681	BG719681.602689891
C	30	6	100.0	31	9	A1118841	A1118841.uc14d10.x
C	31	6	100.0	31	9	A1174159	A1174159.vz84e05.r
C	32	6	100.0	31	9	A1736496	A1736496.sb29d11.y
C	33	6	100.0	31	10	N94283	N94283.za26f01.r1
C	34	6	100.0	31	12	TA106A06P	AL453090.T. bruce1
C	35	6	100.0	32	10	BF168323	BF168323.601774306
C	36	6	100.0	32	12	AZ642287	AZ642287.1M0505D16
C	37	6	100.0	32	12	TA121A02P	AL462540.T. bruce1
C	38	6	100.0	32	12	TA12F010	AL451365.T. bruce1
C	39	6	100.0	33	12	AZ379585	AZ379585.1M0134P16
C	40	6	100.0	34	9	A1225227	A1225227.qx12b04.x
C	41	6	100.0	34	9	A1971896	A1971896.wv29g12.x
C	42	6	100.0	34	10	B1692262	B1692262.603342729
C	43	6	100.0	34	10	BJ055330	BJ055330.BJ055330
C	44	6	100.0	34	12	A0050771	A0050771.nhxb0001C
C	45	6	100.0	34	12	AZ438586	AZ438586.1M0228E15

#### ALIGNMENTS

RESULT 1  
AA911671/c 19 bp mRNA linear EST 10-JUN-1998  
LOCUS  
DEFINITION  
AA911671.1 Soares\_NFL.T\_GBC.S1 Homo sapiens cDNA clone  
IMAGE:1526823.3' similar to TR:018444.018444 COSMID C34D4.  
contains MSRL.b2 MSRL repetitive element ;, mRNA sequence.

ACCESSION  
AA911671  
VERSION  
AA911671.1 GI:3051035  
KEYWORDS  
EST.  
SOURCE  
human.  
ORGANISM  
Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE  
AUTHORS  
NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.  
TITLE  
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
Tumor Gene Index

JOURNAL  
COMMENT  
Unpublished (1997)  
Contact: Robert Strausberg, Ph.D.  
Email: cgaps-remail.nih.gov

This clone is available royalty-free through LNL ; contact the  
IMAGE Consortium (infoimage.lnl.gov) for further information.  
Trace considered overall poor quality  
Insert Length: 682 Std Error: 0.00  
Seq primer: -40m13 fwd. ET from Amersham  
High quality sequence stop: 1.  
Location/Qualifiers

#### FEATURES

source  
1..19  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:1526823"  
/clone\_id="Soares\_NFL.T\_GBC.S1"  
/lab\_host="DH10B"  
/note="Organ: pooled; Vector: pT73D-Pac (Pharmacia) with  
a modified polylinker; Site 1: Not I; Site 2: Eco RI;  
Equal amounts of plasmid DNA from three normalized  
libraries (fetal lung NBHL19W, testis NHT, and B-cell  
NCL1CGAP-GCBI) were mixed, and ss circles were made in  
vitro. Following HAP purification, this DNA was used as  
tracer in a subtractive hybridization reaction. The driver

was PCR-amplified cDNAs from pools of 5,000 clones made from the same 3 libraries. The pools consisted of 1.M.A.G.E. clones 297480-302087, 682632-687239, 726408-728711, and 729096-731399. Subtraction by Bento Soares and M. Fatima Bonaldo.

BASE COUNT 3 a 2 c 11 g 3 t

inducible derivative of plasmid R1. The vector was ligated with adaptors complementary to the insert adaptors and purified. The sheared, adaptor DNA was annealed to adaptor vector DNA, and transformed into chemically competent E. coli XL10-Gold (Stratagene) cells and selected for ampicillin resistance.

Query Match 100.0%; Score 6; DB 9; Length 19;  
Best Local Similarity 100.0%; Pred. No. 5.9e+05;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 ccgttc 6  
111111  
DB 11 ccgttc 6

Query Match 100.0%; Score 6; DB 12; Length 21;  
Best Local Similarity 100.0%; Pred. No. 6e+05;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 ccgttc 6  
111111  
DB 15 ccgttc 10

RESULT 2  
A2347845 21 bp DNA linear GSS 29-SEP-2000  
LOCUS 1M0084106F Mouse 10kb plasmid UUCG1M library Mus musculus genomic  
DEFINITION clone UUCG1M0084106 F, DNA sequence.  
ACCESSION A2347845  
VERSION A2347845.1 GI:10427082  
KEYWORDS GSS.  
SOURCE house mouse.  
ORGANISM Mus musculus

## Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Rodentia; Sclurognathi; Muridae; Murinae; Mus.  
1 (bases 1 to 21)

REFERENCE  
AUTHORS Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamill,C., Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly,M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A. and Wright,D., Weiss,R.

TITLE Mouse whole genome scaffolding with paired end reads from 10kb plasmid inserts

JOURNAL  
COMMENT Unpublished (2000)  
Contact: Robert B. Weiss  
University of Utah Genome Center  
University of Utah

Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT 84112, USA  
Tel: 801 585 5606  
Fax: 801 585 7177  
Email: ddunn@genetics.utah.edu  
Insert Length: 10000 Std Error: 0.00  
Plate: 0084 row: L column: 06  
Seq primer: CGTTGTAACGACGCGCAGT  
Class: plasmid ends  
High quality sequence stop: 21.  
Location/Qualifiers

FEATURES  
Source 1..21

/organism="Mus musculus"  
/strain="C57BL/6J"  
/db\_xref="taxon:10090"  
/clone="UUCG1M0084106"  
/clone.lib="Mouse 10kb plasmid UUCG1M library"  
/sex="Male"

/lab\_host="E. Coli strain XL10-Gold, T1-resistant, F-"  
/note="Vector: PWD42nv; Purified genomic DNA from M. musculus C57BL/6J (male) was obtained from the Jackson Laboratory Mouse DNA Resource  
(http://www.jax.org/resources/documents/dnares/). The DNA was hydrodynamically sheared by repeated passage through a 0.005 inch orifice at constant velocity. The sheared DNA was blunt end-repaired with T4 DNA polymerase and T4 polynucleotide kinase. Adaptor oligonucleotides were ligated to the blunt ends in high molar excess. The adaptor DNA was purified and size-selected for a 9.5 to 10.5 kb range using preparative agarose gel electrophoresis. Vector DNA was prepared from a derivative of PWD42 (91473114|gb|AF129072.1), a copy number

RESULT 3  
A2811866 22 bp DNA linear GSS 20-FEB-2001  
LOCUS 2M0078M08F Mouse 10kb plasmid UUCG1M library Mus musculus genomic  
DEFINITION clone UUCG2M0078M08 F, DNA sequence.  
ACCESSION A2811866  
VERSION A2811866.1 GI:12980548  
KEYWORDS GSS.  
SOURCE house mouse.  
ORGANISM Mus musculus

Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sclurognathi; Muridae; Murinae; Mus.  
1 (bases 1 to 22)

REFERENCE  
AUTHORS Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamill,C., Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly,M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A. and Wright,D., Weiss,R.

TITLE Mouse whole genome scaffolding with paired end reads from 10kb plasmid inserts

JOURNAL  
COMMENT Unpublished (2000)  
Contact: Robert B. Weiss  
University of Utah Genome Center  
University of Utah

Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT 84112, USA  
Tel: 801 585 5606  
Fax: 801 585 7177  
Email: ddunn@genetics.utah.edu  
Insert Length: 10000 Std Error: 0.00  
Plate: 0078 row: M column: 08  
Seq primer: CGTTGTAACGACGCGCAGT  
Class: plasmid ends  
High quality sequence stop: 22.  
Location/Qualifiers

FEATURES  
Source 1..22

/organism="Mus musculus"  
/strain="C57BL/6J"  
/db\_xref="taxon:10090"  
/clone="UUCG2M0078M08"  
/clone.lib="Mouse 10kb plasmid UUCG1M library"  
/sex="Male"

/lab\_host="E. Coli strain XL10-Gold, T1-resistant, F-"  
/note="Vector: PWD42nv; Purified genomic DNA from M. musculus C57BL/6J (male) was obtained from the Jackson Laboratory Mouse DNA Resource  
(http://www.jax.org/resources/documents/dnares/). The DNA was hydrodynamically sheared by repeated passage through a 0.005 inch orifice at constant velocity. The sheared DNA was blunt end-repaired with T4 DNA polymerase and T4 polynucleotide kinase. Adaptor oligonucleotides were ligated to the blunt ends in high molar excess. The adaptor DNA was purified and size-selected for a 9.5 to 10.5 kb range using preparative agarose gel electrophoresis. Vector DNA was prepared from a derivative

of pMD42 (g114732114|gb|AF129072.1), a copy-number inducible derivative of plasmid R1. The vector was ligated with adaptors complementary to the insert adaptors and purified. The sheared, adapted mouse DNA was annealed to adapted vector DNA, and transformed into chemically-competent *E. coli* XL10-Gold (Stratagene) cells and selected for ampicillin resistance."

BASE COUNT 8 a 5 c 7 g 2 t

ORIGIN

Query Match 100.0%; Score 6; DB 12; Length 22;  
Best Local Similarity 100.0%; Pred. No. 6e+05;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ccgttc 6  
111111

Db 16 CCGTTC 11

#### RESULT 4

AZ653869/c

LOCUS 23 bp DNA linear GSS 14-DEC-2000  
DEFINITION IM0527D14R Mouse 10kb plasmid UUGC1M library Mus musculus genomic

ACCESSION AZ653869

VERSION AZ653869.1 GI:11791015

KEYWORDS GSS.

SOURCE house mouse.

ORGANISM Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.

REFERENCE 1 (bases 1 to 23)

Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamll,C., Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly,M., Rose,M., Rose,R., Stokes,R., Tinney,A., von Niederhausen,A. and Wright,D., Weiss,R.  
Mouse whole genome scaffolding with paired end reads from 10kb plasmid inserts

TITLE

JOURNAL

COMMENT

Unpublished (2000)

Contact: Robert B. Weiss

University of Utah Genome Center

Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT

84112, USA

Tel: 801 585 5606

Fax: 801 585 7177

Email: ddunn@genetics.utah.edu

Insert Length: 10000 Std Error: 0.00

Plate: 0527 row: D column: 14

Seq primer: CACACAGGAAACACGTATGACC

Class: plasmid ends

High quality sequence stop: 23.

Location/Qualifiers

#### FEATURES

source

1. .23

/organism="Mus musculus"

/strain="C57BL/6J"

/db\_xref="taxon:10090"

/clone="UUGC1M0527D14"

/clone\_lib="Mouse 10kb plasmid UUGC1M library"

/sex="Male"

/lab\_host="E. Coli strain XL10-Gold, T1-resistant, F-"

/note="Vector: PMD42nv; Purified genomic DNA from M.

musculus C57BL/6J (male) was obtained from the Jackson

Laboratory Mouse DNA Resource

(http://www.jax.org/resources/documents/dnares/). The DNA

was hydrodynamically sheared by repeated passage through a

0.005 inch orifice at constant velocity. The sheared DNA

was blunt end-repaired with T4 DNA polymerase and T4

polynucleotide kinase. Adaptor oligonucleotides were

ligated to the blunt ends in high molar excess. The

adapted DNA was purified and size-selected for a 9.5 to

10.5 kb range using preparative agarose gel

electrophoresis. Vector DNA was prepared from a derivative of pMD42 (g114732114|gb|AF129072.1), a copy-number inducible derivative of plasmid R1. The vector was ligated with adaptors complementary to the insert adaptors and purified. The sheared, adapted mouse DNA was annealed to adapted vector DNA, and transformed into chemically-competent *E. coli* XL10-Gold (Stratagene) cells and selected for ampicillin resistance."

BASE COUNT 4 a 1 c 13 g 5 t

ORIGIN

Query Match 100.0%; Score 6; DB 12; Length 23;  
Best Local Similarity 100.0%; Pred. No. 6.1e+05;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ccgttc 6  
111111

Db 11 CCGTTC 6

#### RESULT 5

AZ316663

LOCUS 24 bp DNA linear GSS 29-SEP-2000  
DEFINITION IM0034G22R Mouse 10kb plasmid UUGC1M library Mus musculus genomic

ACCESSION AZ316663

VERSION AZ316663.1 GI:10364703

KEYWORDS GSS.

SOURCE house mouse.

ORGANISM Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.

REFERENCE 1 (bases 1 to 24)

Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamll,C., Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly,M., Rose,M., Rose,R., Stokes,R., Tinney,A., von Niederhausen,A. and Wright,D., Weiss,R.  
Mouse whole genome scaffolding with paired end reads from 10kb plasmid inserts

TITLE

JOURNAL

COMMENT

Unpublished (2000)

Contact: Robert B. Weiss

University of Utah Genome Center

Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT

84112, USA

Tel: 801 585 5606

Fax: 801 585 7177

Email: ddunn@genetics.utah.edu

Insert Length: 10000 Std Error: 0.00

Plate: 0034 row: G column: 22

Seq primer: CACACAGGAAACACGTATGACC

Class: plasmid ends

High quality sequence stop: 24.

Location/Qualifiers

#### FEATURES

source

1. .24

/organism="Mus musculus"

/strain="C57BL/6J"

/db\_xref="taxon:10090"

/clone="UUGC1M0034G22"

/clone\_lib="Mouse 10kb plasmid UUGC1M library"

/sex="Male"

/lab\_host="E. Coli strain XL10-Gold, T1-resistant, F-"

/note="Vector: PMD42nv; Purified genomic DNA from M.

musculus C57BL/6J (male) was obtained from the Jackson

Laboratory Mouse DNA Resource

(http://www.jax.org/resources/documents/dnares/). The DNA

was hydrodynamically sheared by repeated passage through a

0.005 inch orifice at constant velocity. The sheared DNA

was blunt end-repaired with T4 DNA polymerase and T4

polynucleotide kinase. Adaptor oligonucleotides were

ligated to the blunt ends in high molar excess. The

adapted DNA was purified and size-selected for a 9.5 to

10.5 kb range using preparative agarose gel electrophoresis. Vector DNA was prepared from a derivative of pMD42 (g114732114|gb|AF129072.1), a copy-number inducible derivative of plasmid R1. The vector was ligated with adaptors complementary to the insert adaptors and purified. The sheared, adaptor mouse DNA was annealed to adapted vector DNA, and transformed into chemically-competent *E. coli* XL10-Gold (Stratagene) cells and selected for ampicillin resistance."

BASE COUNT 5 a 9 c 4 g 6 t

ORIGIN

Query Match 100.0%; Score 6; DB 12; Length 24;  
Best Local Similarity 100.0%; Pred. No. 6.1e+05;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ccgttc 6  
|||||  
Db 9 CCGTTC 14

RESULT 6 A2436588 24 bp DNA linear GSS 03-OCT-2000  
LOCUS 1M0224H20F Mouse 10kb plasmid UUGC1M library Mus musculus genomic  
DEFINITION clone UUGC1M0224H20 F, DNA sequence.

ACCESSION A2436588  
VERSION A2436588.1 GI:10560601

KEYWORDS GSS.  
SOURCE house mouse.  
ORGANISM Mus musculus

REFERENCE 1 (bases 1 to 24)  
AUTHORS Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C., Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly,M., Rose,M., Rose,R., Stokes,R., Tinney,A., von Niederhausern,A. and Wright,D., Weiss,R.

TITLE Mouse whole genome scaffolding with paired end reads from 10kb plasmid inserts

JOURNAL Unpublished (2000)  
COMMENT Contact: Robert B. Weiss  
University of Utah Genome Center  
University of Utah  
Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLIC, UT  
84112, USA  
Tel: 801 585 5606  
Fax: 801 585 7177  
Email: ddunn@genetics.utah.edu

Insert length: 10000 Std Error: 0.00  
Plate: 0224 row: H column: 20  
Seq primer: CGTGTAAACGACGCCAGT  
Class: plasmid ends  
High quality sequence stop: 24.

FEATURES  
source

1. .24 Location/Qualifiers

/organism="Mus musculus"  
/strain="C57BL/6J"  
/db\_xref="taxon:10090"  
/clone="UUGC1M0224H20"  
/clone\_1lb="Mouse 10kb plasmid UUGC1M library"  
/sex="Male"  
/lab\_host="E. Coli strain XL10-Gold, T1-resistant, F-"  
/note="Vector: pMD42nv; Purified genomic DNA from M. musculus C57BL/6J (male) was obtained from the Jackson Laboratory Mouse DNA Resource (http://www.jax.org/resources/documents/dnares/). The DNA was hydrodynamically sheared by repeated passage through a 0.005 inch orifice at constant velocity. The sheared DNA was blunt end-repaired with T4 DNA polymerase and T4 polynucleotide kinase. Adaptor oligonucleotides were ligated to the blunt ends in high molar excess. The

adapted DNA was purified and size-selected for a 9.5 to 10.5 kb range using preparative agarose gel electrophoresis. Vector DNA was prepared from a derivative of pMD42 (g114732114|gb|AF129072.1), a copy-number inducible derivative of plasmid R1. The vector was ligated with adaptors complementary to the insert adaptors and purified. The sheared, adaptor mouse DNA was annealed to adapted vector DNA, and transformed into chemically-competent *E. coli* XL10-Gold (Stratagene) cells and selected for ampicillin resistance."

BASE COUNT 3 a 5 c 13 g 3 t

ORIGIN

Query Match 100.0%; Score 6; DB 12; Length 24;  
Best Local Similarity 100.0%; Pred. No. 6.1e+05;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ccgttc 6  
|||||  
Db 12 CCGTTC 7

RESULT 7 A2446206 24 bp DNA linear GSS 04-OCT-2000  
LOCUS 1M0242I06R Mouse 10kb plasmid UUGC1M library Mus musculus genomic  
DEFINITION clone UUGC1M0242I06 R, DNA sequence.

ACCESSION A2446206  
VERSION A2446206.1 GI:10596787

KEYWORDS GSS.  
SOURCE house mouse.  
ORGANISM Mus musculus

REFERENCE 1 (bases 1 to 24)  
AUTHORS Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C., Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly,M., Rose,M., Rose,R., Stokes,R., Tinney,A., von Niederhausern,A. and Wright,D., Weiss,R.

TITLE Mouse whole genome scaffolding with paired end reads from 10kb plasmid inserts

JOURNAL Unpublished (2000)  
COMMENT Contact: Robert B. Weiss  
University of Utah Genome Center  
University of Utah  
Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLIC, UT  
84112, USA  
Tel: 801 585 5606  
Fax: 801 585 7177  
Email: ddunn@genetics.utah.edu

Insert length: 10000 Std Error: 0.00  
Plate: 0242 row: I column: 06  
Seq primer: CACACAGCAACACGTATGACC  
Class: plasmid ends  
High quality sequence stop: 24.

FEATURES  
source

1. .24 Location/Qualifiers

/organism="Mus musculus"  
/strain="C57BL/6J"  
/db\_xref="taxon:10090"  
/clone="UUGC1M0242I06"  
/clone\_1lb="Mouse 10kb plasmid UUGC1M library"  
/sex="Male"  
/lab\_host="E. Coli strain XL10-Gold, T1-resistant, F-"  
/note="Vector: pMD42nv; Purified genomic DNA from M. musculus C57BL/6J (male) was obtained from the Jackson Laboratory Mouse DNA Resource (http://www.jax.org/resources/documents/dnares/). The DNA was hydrodynamically sheared by repeated passage through a 0.005 inch orifice at constant velocity. The sheared DNA was blunt end-repaired with T4 DNA polymerase and T4 polynucleotide kinase. Adaptor oligonucleotides were

ligated to the blunt ends in high molar excess. The adapted DNA was purified and size-selected for a 9.5 to 10.5 kb range using preparative agarose gel electrophoresis. Vector DNA was prepared from a derivative of pMD42 (g114732114|gblAF129072.1), a copy-number inducible derivative of plasmid R1. The vector was ligated with adaptors complementary to the insert adaptors and purified. The sheared, adapted mouse DNA was annealed to adapted vector DNA, and transformed into chemically-competent E. coli XL10-Gold (Stratagene) cells and selected for ampicillin resistance."

BASE COUNT 4 a 6 c 10 g 4 t

ORIGIN

Query Match 100.0%; Score 6; DB 12; Length 24;  
Best Local Similarity 100.0%; Pred. No. 6.1e+05;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1 ccgttc 6  
|||||  
Db 20 CCGTTC 15

RESULT 8  
A2662500/c 24 bp DNA linear GSS 14-DEC-2000

LOCUS 1M0541G07R Mouse 10kb plasmid UUGC1M library Mus musculus genomic  
DEFINITION clone UUGC1M0541G07 R, DNA sequence.

ACCESSION A2662500

VERSION A2662500.1 GI:11799646

KEYWORDS GSS.

SOURCE house mouse.

ORGANISM Mus musculus

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus. 1 (bases 1 to 24)

AUTHORS Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamill,C., Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly,M., Rose,M., Rose,R., Stokes,R., Tinney,A., von Niederhausern,A. and Wright,D.,Weiss,R.

TITLE Mouse whole genome scaffolding with paired end reads from 10kb plasmid inserts

JOURNAL Unpublished (2000)

COMMENT Contact: Robert B. Weiss

University of Utah Genome Center

Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLIC, UT 84112, USA

Tel: 801 585 5606

Fax: 801 585 7177

Email: ddunn@genetics.utah.edu

Insert Length: 10000 Std Error: 0.00

Plate: 0541 row: G column: 07

Seq primer: CACACAGCAACACATGACAC

Class: plasmid ends

High quality sequence stop: 24.

Location/Qualifiers

1. 24

/organism="Mus musculus"

/strain="C57BL/6J"

/db\_xref="taxon:10090"

/clone="UUGC1M0541G07"

/clone\_lib="Mouse 10kb plasmid UUGC1M library"

/sex="Male"

/lab\_host="E. Coli strain XL10-Gold, T1-resistant, F-"

/note="Vector: PMD42nv; Purified genomic DNA from M. musculus C57BL/6J (male) was obtained from the Jackson Laboratory Mouse DNA Resource

(http://www.jax.org/resources/documents/dnares/). The DNA was hydrodynamically sheared by repeated passage through a

0.005 inch orifice at constant velocity. The sheared DNA was blunt end-repaired with T4 DNA polymerase and T4

polynucleotide kinase. Adaptor oligonucleotides were ligated to the blunt ends in high molar excess. The adapted DNA was purified and size-selected for a 9.5 to 10.5 kb range using preparative agarose gel electrophoresis. Vector DNA was prepared from a derivative of pMD42 (g114732114|gblAF129072.1), a copy-number inducible derivative of plasmid R1. The vector was ligated with adaptors complementary to the insert adaptors and purified. The sheared, adapted mouse DNA was annealed to adapted vector DNA, and transformed into chemically-competent E. coli XL10-Gold (Stratagene) cells and selected for ampicillin resistance."

BASE COUNT 8 a 3 c 7 g 6 t

ORIGIN

Query Match 100.0%; Score 6; DB 12; Length 24;  
Best Local Similarity 100.0%; Pred. No. 6.1e+05;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1 ccgttc 6  
|||||  
Db 22 CCGTTC 17

RESULT 9

A2789936 24 bp DNA linear GSS 16-FEB-2001

LOCUS 2M0038L17F Mouse 10kb plasmid UUGC1M library Mus musculus genomic

DEFINITION clone UUGC2M0038L17 F, DNA sequence.

ACCESSION A2789936

VERSION A2789936.1 GI:12931470

KEYWORDS GSS.

SOURCE house mouse.

ORGANISM Mus musculus

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus. 1 (bases 1 to 24)

AUTHORS Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamill,C., Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly,M., Rose,M., Rose,R., Stokes,R., Tinney,A., von Niederhausern,A. and Wright,D.,Weiss,R.

TITLE Mouse whole genome scaffolding with paired end reads from 10kb plasmid inserts

JOURNAL Unpublished (2000)

COMMENT Contact: Robert B. Weiss

University of Utah Genome Center

Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLIC, UT 84112, USA

Tel: 801 585 5606

Fax: 801 585 7177

Email: ddunn@genetics.utah.edu

Insert Length: 10000 Std Error: 0.00

Plate: 0038 row: L column: 17

Seq primer: CGTGTAAACGACGCCACAT

Class: plasmid ends

High quality sequence stop: 24.

Location/Qualifiers

1. 24

/organism="Mus musculus"

/strain="C57BL/6J"

/db\_xref="taxon:10090"

/clone="UUGC2M0038L17"

/clone\_lib="Mouse 10kb plasmid UUGC1M library"

/sex="Male"

/lab\_host="E. Coli strain XL10-Gold, T1-resistant, F-"

/note="Vector: PMD42nv; Purified genomic DNA from M. musculus C57BL/6J (male) was obtained from the Jackson Laboratory Mouse DNA Resource

(http://www.jax.org/resources/documents/dnares/). The DNA was hydrodynamically sheared by repeated passage through a

0.005 inch orifice at constant velocity. The sheared DNA



DEFINITION IM0236K13R Mouse 10kb plasmid UUGC1M library Mus musculus genomic  
clone UUGC1M0236K13 R, DNA sequence.

ACCESSION A2442576  
VERSION A2442576.1 GI:10589722  
KEYWORDS GSS.

SOURCE house mouse.  
ORGANISM Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
1 (bases 1 to 25)

REFERENCE 1  
AUTHORS Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C.,  
Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly,  
'M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A.  
and Wright,D., Weiss,R.  
Mouse whole genome scaffolding with paired end reads from 10kb  
plasmid inserts  
Unpublished (2000)  
JOURNAL Contact: Robert B. Weiss  
COMMENT University of Utah  
Genome Center  
Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLG, UT  
84112, USA  
Tel: 801 585 5606  
Fax: 801 585 7177  
Email: ddunn@genetics.utah.edu  
Insert Length: 10000 Std Error: 0.00  
Plate: 0236 row: K column: 13  
Seq primer: CACACAGAAACACGTATGACC  
Class: plasmid ends  
High quality sequence stop: 25.  
Location/Qualifiers  
1. 25  
/organism="Mus musculus"  
/strain="C57BL/6J"  
/db\_xref="taxon:10090"  
/clone="UUGC1M0236K13"  
/clone\_lib="Mouse 10kb plasmid UUGC1M library"  
/sex="Male"  
/lab\_host="E. Coli strain XL10-Gold, T1-resistant, F-"  
/note="Vector: PWD42nv; Purified genomic DNA from M.  
musculus C57BL/6J (male) was obtained from the Jackson  
Laboratory Mouse DNA Resource  
(http://www.jax.org/resources/documents/dnares/). The DNA  
was hydrodynamically sheared by repeated passage through a  
0.005 inch orifice at constant velocity. The sheared DNA  
was blunt end-repaired with T4 DNA polymerase and T4  
polynucleotide kinase. Adaptor oligonucleotides were  
ligated to the blunt ends in high molar excess. The  
adapted DNA was purified and size-selected for a 9.5 to  
10.5 kb range using preparative agarose gel  
electrophoresis. Vector DNA was prepared from a derivative  
of PWD42 (gii4732114|gb|AF129072.1), a copy-number  
inducible derivative of plasmid R1. The vector was ligated  
with adaptors complementary to the insert adaptors and  
purified. The sheared, adapted mouse DNA was annealed to  
adapted vector DNA, and transformed into  
chemically-competent E. coli XL10-Gold (Stratagene) cells  
and selected for ampicillin resistance."

BASE COUNT 9 a 6 c 6 g 4 t  
ORIGIN

Query Match 100.0%; Score 6; DB 12; Length 25;  
Best Local Similarity 100.0%; Pred. No. 6.1e+05;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1 ccgttc 6  
|||||  
Db 9 CCGTTC 4

RESULT 13  
A2621312/c

LOCUS A2621312 25 bp DNA linear GSS 13-DEC-2000  
DEFINITION IM0454P19F Mouse 10kb plasmid UUGC1M library Mus musculus genomic  
clone UUGC1M0454P19 F, DNA sequence.

ACCESSION A2621312  
VERSION A2621312.1 GI:11743502  
KEYWORDS GSS.

SOURCE house mouse.  
ORGANISM Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
1 (bases 1 to 25)

REFERENCE 1  
AUTHORS Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C.,  
Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly,  
'M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A.  
and Wright,D., Weiss,R.  
Mouse whole genome scaffolding with paired end reads from 10kb  
plasmid inserts  
Unpublished (2000)  
JOURNAL Contact: Robert B. Weiss  
COMMENT University of Utah  
Genome Center  
Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLG, UT  
84112, USA  
Tel: 801 585 5606  
Fax: 801 585 7177  
Email: ddunn@genetics.utah.edu  
Insert Length: 10000 Std Error: 0.00  
Plate: 0454 row: P column: 19  
Seq primer: CGTGTAAACGACGCCAGT  
Class: plasmid ends  
High quality sequence stop: 25.  
Location/Qualifiers  
1. 25  
/organism="Mus musculus"  
/strain="C57BL/6J"  
/db\_xref="taxon:10090"  
/clone="UUGC1M0454P19"  
/clone\_lib="Mouse 10kb plasmid UUGC1M library"  
/sex="Male"  
/lab\_host="E. Coli strain XL10-Gold, T1-resistant, F-"  
/note="Vector: PWD42nv; Purified genomic DNA from M.  
musculus C57BL/6J (male) was obtained from the Jackson  
Laboratory Mouse DNA Resource  
(http://www.jax.org/resources/documents/dnares/). The DNA  
was hydrodynamically sheared by repeated passage through a  
0.005 inch orifice at constant velocity. The sheared DNA  
was blunt end-repaired with T4 DNA polymerase and T4  
polynucleotide kinase. Adaptor oligonucleotides were  
ligated to the blunt ends in high molar excess. The  
adapted DNA was purified and size-selected for a 9.5 to  
10.5 kb range using preparative agarose gel  
electrophoresis. Vector DNA was prepared from a derivative  
of PWD42 (gii4732114|gb|AF129072.1), a copy-number  
inducible derivative of plasmid R1. The vector was ligated  
with adaptors complementary to the insert adaptors and  
purified. The sheared, adapted mouse DNA was annealed to  
adapted vector DNA, and transformed into  
chemically-competent E. coli XL10-Gold (Stratagene) cells  
and selected for ampicillin resistance."

BASE COUNT 4 a 2 c 15 g 4 t  
ORIGIN

Query Match 100.0%; Score 6; DB 12; Length 25;  
Best Local Similarity 100.0%; Pred. No. 6.1e+05;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1 ccgttc 6  
|||||  
Db 15 CCGTTC 10

RESULT 14

AZ372925/c  
LOCUS AZ372925 26 bp DNA linear GSS 02-OCT-2000  
DEFINITION M0125D04F Mouse 10kb plasmid UUGC1M library Mus musculus genomic  
ACCESSION AZ372925  
VERSION AZ372925.1 GI:10486625  
KEYWORDS GSS.  
SOURCE house mouse.  
ORGANISM Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
1 (bases 1 to 26)  
Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamill,C.,  
Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly  
,M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A.  
and Wright,D., Weiss,R.  
Mouse whole genome scaffolding with paired end reads from 10kb  
plasmid inserts  
TITLE  
JOURNAL  
COMMENT Unpublished (2000)  
Contact: Robert B. Weiss  
University of Utah Genome Center  
Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT  
84112, USA  
Tel: 801 585 5606  
Fax: 801 585 7177  
Email: ddunn@genetics.utah.edu  
Insert Length: 10000 Std Error: 0.00  
Plate: 0125 row: D column: 04  
Seq primer: CGTGTAAACGACGCGCAGT  
Class: plasmid ends  
High quality sequence stop: 26.  
Location/Qualifiers  
1..26  
/organism="Mus musculus"  
/strain="C57BL/6J"  
/db\_xref="taxon:10090"  
/clone="UUGC1M0125D04"  
/clone\_1lb="Mouse 10kb plasmid UUGC1M library"  
/sex="Male"  
/lab\_host="E. Coli strain XL10-Gold, T1-resistant, F-"  
/note="Vector: PMD42nv; Purified genomic DNA from M.  
musculus C57BL/6J (male) was obtained from the Jackson  
Laboratory Mouse DNA Resource  
(http://www.jax.org/resources/documents/dnares/). The DNA  
was hydrodynamically sheared by repeated passage through a  
0.005 inch orifice at constant velocity. The sheared DNA  
was blunt end-repaired with T4 DNA polymerase and T4  
polynucleotide kinase. Adaptor oligonucleotides were  
ligated to the blunt ends in high molar excess. The  
adaptor DNA was purified and size-selected for a 9.5 to  
10.5 kb range using preparative agarose gel  
electrophoresis. Vector DNA was prepared from a derivative  
of pMD42 (g11473211419b1AF129072.1), a copy number  
inducible derivative of plasmid R1. The vector was ligated  
with adaptors complementary to the insert adaptors and  
purified. The sheared, adaptor mouse DNA was annealed to  
adaptor vector DNA, and transformed into  
chemically-competent E. coli XL10-Gold (Stratagene) cells  
and selected for ampicillin resistance."

BASE COUNT 8 a 4 c 9 g 5 t  
ORIGIN

Query Match 100.0%; Score 6; DB 12; Length 26;  
Best Local Similarity 100.0%; Pred. No. 6.2e+05;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ccgttc 6  
|||||  
DB 16 CCGTTC 11

RESULT 15  
AZ819947  
LOCUS AZ819947 26 bp DNA linear GSS 20-FEB-2001  
DEFINITION 2M0091E20R Mouse 10kb plasmid UUGC1M library Mus musculus genomic  
ACCESSION AZ819947  
VERSION AZ819947.1 GI:12989855  
KEYWORDS GSS.  
SOURCE house mouse.  
ORGANISM Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
1 (bases 1 to 26)  
Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamill,C.,  
Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly  
,M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A.  
and Wright,D., Weiss,R.  
Mouse whole genome scaffolding with paired end reads from 10kb  
plasmid inserts  
TITLE  
JOURNAL  
COMMENT Unpublished (2000)  
Contact: Robert B. Weiss  
University of Utah Genome Center  
Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT  
84112, USA  
Tel: 801 585 5606  
Fax: 801 585 7177  
Email: ddunn@genetics.utah.edu  
Insert Length: 10000 Std Error: 0.00  
Plate: 0091 row: E column: 20  
Seq primer: CACACGAGAAACGCTGACAC  
Class: plasmid ends  
High quality sequence stop: 26.  
Location/Qualifiers  
1..26  
/organism="Mus musculus"  
/strain="C57BL/6J"  
/db\_xref="taxon:10090"  
/clone="UUGC2M0091E20"  
/clone\_1lb="Mouse 10kb plasmid UUGC1M library"  
/sex="Male"  
/lab\_host="E. Coli strain XL10-Gold, T1-resistant, F-"  
/note="Vector: PMD42nv; Purified genomic DNA from M.  
musculus C57BL/6J (male) was obtained from the Jackson  
Laboratory Mouse DNA Resource  
(http://www.jax.org/resources/documents/dnares/). The DNA  
was hydrodynamically sheared by repeated passage through a  
0.005 inch orifice at constant velocity. The sheared DNA  
was blunt end-repaired with T4 DNA polymerase and T4  
polynucleotide kinase. Adaptor oligonucleotides were  
ligated to the blunt ends in high molar excess. The  
adaptor DNA was purified and size-selected for a 9.5 to  
10.5 kb range using preparative agarose gel  
electrophoresis. Vector DNA was prepared from a derivative  
of pMD42 (g11473211419b1AF129072.1), a copy number  
inducible derivative of plasmid R1. The vector was ligated  
with adaptors complementary to the insert adaptors and  
purified. The sheared, adaptor mouse DNA was annealed to  
adaptor vector DNA, and transformed into  
chemically-competent E. coli XL10-Gold (Stratagene) cells  
and selected for ampicillin resistance."

BASE COUNT 3 a 9 c 4 g 10 t  
ORIGIN

Query Match 100.0%; Score 6; DB 12; Length 26;  
Best Local Similarity 100.0%; Pred. No. 6.2e+05;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ccgttc 6  
|||||  
DB 13 CCGTTC 18



Tue Jul 30 09:10:32 2002

us-09-530-663b-16.rst

Page 9

Search completed: July 29, 2002, 23:22:44  
Job time: 6820 sec

---



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 29, 2002, 23:56:13 ; Search time 65.09 Seconds  
(without alignments)  
22.643 Million cell updates/sec

Title: US-09-530-663B-16

Perfect score: 6

Sequence: 1 ccgttc 6

Scoring table: IDENTITY\_NUC

Searched: 383533 seqs, 122816752 residues

Total number of hits satisfying chosen parameters: 767066

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued\_Patents\_NA.\*  
1: /cgn2\_6/ptodata/2/ina/5A.COMB.seq:\*  
2: /cgn2\_6/ptodata/2/ina/5B.COMB.seq:\*  
3: /cgn2\_6/ptodata/2/ina/6A.COMB.seq:\*  
4: /cgn2\_6/ptodata/2/ina/6B.COMB.seq:\*  
5: /cgn2\_6/ptodata/2/ina/PT05.COMB.seq:\*  
6: /cgn2\_6/ptodata/2/ina/backfiles1.seq:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	6	100.0	9	2	US-08-642-045B-2
2	6	100.0	9	3	US-08-852-268-4
3	6	100.0	10	2	US-08-545-253A-8
4	6	100.0	10	3	US-08-719-337-8
5	6	100.0	10	4	US-08-878-835A-7
6	6	100.0	12	1	US-07-974-447-10
7	6	100.0	12	1	US-08-149-199-10
8	6	100.0	12	1	US-08-411-727-5
9	6	100.0	12	1	US-08-858-767-6
10	6	100.0	12	2	US-08-858-767-8
11	6	100.0	12	2	US-08-863-028-6
12	6	100.0	12	2	US-08-863-028-6
13	6	100.0	12	3	US-09-115-061-10
14	6	100.0	12	4	US-09-261-079-10
15	6	100.0	14	1	US-08-424-921-5
16	6	100.0	14	1	US-08-424-921-6
17	6	100.0	14	1	US-08-663-769-11
18	6	100.0	14	1	US-08-193-627-2
19	6	100.0	14	3	US-08-985-162-1835
20	6	100.0	14	3	US-08-985-162-1835
21	6	100.0	15	1	US-08-182-968A-26
22	6	100.0	15	1	US-08-182-968A-26
23	6	100.0	15	2	US-08-774-306A-26
24	6	100.0	15	2	US-08-774-306A-26
25	6	100.0	15	3	US-09-064-156A-26
26	6	100.0	15	3	US-09-064-156A-26
27	6	100.0	15	3	US-09-064-156A-26

C 28	6	100.0	15	3	US-08-461-366A-4	Sequence 4, Appl
C 29	6	100.0	15	3	US-08-481-341-14	Sequence 14, Appl
C 30	6	100.0	15	4	US-09-402-764B-11	Sequence 11, Appl
C 31	6	100.0	15	4	US-08-464-514-6	Sequence 6, Appl
C 32	6	100.0	15	4	US-08-486-403-6	Sequence 6, Appl
C 33	6	100.0	15	4	US-09-078-954-17	Sequence 17, Appl
C 34	6	100.0	15	4	US-09-054-837-26	Sequence 26, Appl
C 35	6	100.0	15	5	PCT-US94-00265-14	Sequence 14, Appl
C 36	6	100.0	15	6	5182195-49	Patent No. 5182195
C 37	6	100.0	16	1	US-08-513-841-14	Sequence 14, Appl
C 38	6	100.0	16	2	US-08-696-834-15	Sequence 15, Appl
C 39	6	100.0	16	2	US-08-942-673-14	Sequence 14, Appl
C 40	6	100.0	16	4	US-09-118-317-14	Sequence 14, Appl
C 41	6	100.0	16	4	US-09-134-607A-3	Sequence 3, Appl
C 42	6	100.0	16	4	US-08-679-645-521	Sequence 521, App
C 43	6	100.0	17	1	US-07-879-647A-43	Sequence 43, Appl
C 44	6	100.0	17	1	US-07-879-584A-43	Sequence 43, Appl
C 45	6	100.0	17	1	US-07-879-470A-43	Sequence 43, Appl

## ALIGNMENTS

RESULT 1  
US-08-642-045B-2/c  
; Sequence 2, Application US/08642045B  
; Patent No. 5851804  
; GENERAL INFORMATION:  
; APPLICANT: Snyder, Linda A.  
; APPLICANT: Satishchandran, C.  
; TITLE OF INVENTION: CHIMERIC KANAMYCIN RESISTANCE GENE  
; NUMBER OF SEQUENCES: 25  
; CORRESPONDENCE ADDRESS:  
; ADDRESS: Woodcock Washburn Kurtz Mackiewicz & No. 5851804r1s  
; STREET: One Liberty Place, 46th floor  
; CITY: Philadelphia  
; STATE: Pennsylvania  
; COUNTRY: USA  
; ZIP: 19103  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: WINDOWS  
; SOFTWARE: Wordperfect 6.0/6.1  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/642,045B  
; FILING DATE: 06-MAY-1996  
; CLASSIFICATION: 514  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Deluca, Mark  
; REGISTRATION NUMBER: 33,229  
; REFERENCE/DOCKET NUMBER: APOL-0262  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 215-568-3100  
; TELEFAX: 215-568-3439  
; INFORMATION FOR SEQ ID NO: 2:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 9 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: both  
; TOPOLOGY: both  
; MOLECULE TYPE: DNA  
; US-08-642-045B-2

Query Match 100.0%; Score 6; DB 2; Length 9;  
Best Local Similarity 100.0%; Pred. No. 2.5e+07;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ccgttc 6  
Db 8 CCGTTC 3

RESULT 2  
US-08-852-268-4/C  
; Sequence 4, Application US/08852268  
; Patent No. 6143527  
; GENERAL INFORMATION:  
; APPLICANT: Pachuk, Catherine J.  
; APPLICANT: Samuel, Manoj  
; APPLICANT: Zurawski, John A.  
; APPLICANT: Satishchandran, C.  
; TITLE OF INVENTION: CHAIN REACTION CLONING  
; NUMBER OF SEQUENCES: 21  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Woodcock Washburn Kurtz Mackiewicz & No. 6143527rls  
; STREET: One Liberty Place, 46th floor  
; CITY: Philadelphia  
; STATE: Pennsylvania  
; COUNTRY: USA  
; ZIP: 19103  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: WINDOWS  
; SOFTWARE: Wordperfect  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/852,268  
; FILING DATE:  
; CLASSIFICATION: 435  
; PRIORITY APPLICATION DATA:  
; APPLICATION NUMBER: US 08/642,045  
; FILING DATE: 06-MAY-1996  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Deluda, Mark  
; REGISTRATION NUMBER: 33,229  
; REFERENCE/DOCKET NUMBER: APOL-0265  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 215-568-3100  
; TELEFAX: 215-568-3439  
; INFORMATION FOR SEQ ID NO: 4:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 9 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: both  
; TOPOLOGY: both  
; MOLECULE TYPE: DNA  
; US-08-852-268-4

Query Match 100.0%; Score 6; DB 3; Length 9;  
Best Local Similarity 100.0%; Pred. No. 2.5e+07;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ccgttc 6  
|111111  
Db 8 CCGTTC 3

RESULT 3  
US-08-545-253A-8/C  
; Sequence 8, Application US/08545253A  
; Patent No. 5908978  
; GENERAL INFORMATION:  
; APPLICANT: O'Malley, David M.  
; APPLICANT: Sederoff, Ronald R.  
; APPLICANT: Grattapaglia, Dario  
; APPLICANT: Henry V. Amerson  
; APPLICANT: Phillip Wilcox  
; APPLICANT: E. George Kuhlman  
; TITLE OF INVENTION: METHODS FOR WITHIN FAMILY  
; TITLE OF INVENTION: SELECTION IN  
; TITLE OF INVENTION: WOODY PERENNIALS USING GENETIC MARKERS  
; NUMBER OF SEQUENCES: 26

CORRESPONDENCE ADDRESS:  
ADDRESSEE: Kenneth D. Sibley  
STREET: Post Office Drawer 34009  
CITY: Charlotte  
STATE: No. 5908978th Carolina  
COUNTRY: U.S.A.  
ZIP: 28234  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/545,253A  
; FILING DATE:  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Sibley, Kenneth D.  
; REGISTRATION NUMBER: 31,665  
; REFERENCE/DOCKET NUMBER: 5051-281  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (919) 881-3140  
; TELEFAX: (919) 881-3175  
; TELEX: 575102  
; INFORMATION FOR SEQ ID NO: 8:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 10 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: CDNA  
; US-08-545-253A-8

Query Match 100.0%; Score 6; DB 2; Length 10;  
Best Local Similarity 100.0%; Pred. No. 1.6e+04;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ccgttc 6  
|111111  
Db 6 CCGTTC 1

RESULT 4  
US-08-719-337-8/C  
; Sequence 8, Application US/08719337  
; Patent No. 6054634  
; GENERAL INFORMATION:  
; APPLICANT: O'Malley, David M.  
; APPLICANT: Sederoff, Ronald R.  
; APPLICANT: Grattapaglia, Dario  
; TITLE OF INVENTION: METHODS FOR WITHIN FAMILY SELECTION IN  
; TITLE OF INVENTION: WOODY PERENNIALS USING GENETIC MARKERS  
; NUMBER OF SEQUENCES: 26  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Kenneth D. Sibley  
; STREET: Post Office Drawer 34009  
; CITY: Charlotte  
; STATE: No. 6054634th Carolina  
; COUNTRY: U.S.A.  
; ZIP: 28234  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/719,337  
; FILING DATE: 25-SEP-1996  
; CLASSIFICATION: 047  
; PRIORITY APPLICATION DATA:  
; APPLICATION NUMBER: 08/184,567  
; FILING DATE: 21-JAN-1994

```
ATTORNEY/AGENT INFORMATION:
: NAME: Sibley, Kenneth D.
: REGISTRATION NUMBER: 31,665
: REFERENCE/DOCKET NUMBER: 5051-247
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (919) 881-3140
: TELEFAX: (919) 881-3175
: TELEX: 575102
: INFORMATION FOR SEQ ID NO: 8:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 10 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: MOLECULE TYPE: cDNA
: US-08-719-337-8

Query Match          100.0%; Score 6; DB 3; Length 10;
Best Local Similarity 100.0%; Pred. No. 1.6e+04;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ccgttc 6
   |||||
Db 6 CCGTTC 1

RESULT 5
US-08-878-835A-7/c
: Sequence 7, Application US/08878835A
: Patent No. 6337071
: GENERAL INFORMATION:
: APPLICANT: William Mitchell Molyneux
: TITLE OF INVENTION: Mosquito and/or Flea Control
: NUMBER OF SEQUENCES: 20
: CORRESPONDENCE ADDRESSES:
: ADDRESSEE: D. Peter Hochberg Co., L.P.A.
: STREET: The Baker Building - Sixth Floor 1940 East 6th Street
: STATE: Ohio
: COUNTRY: U.S.A.
: ZIP: 44114
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Diskette, 3.5 inch, 1.44 Mbyte storage
: COMPUTER: IBM Compatible w/ Pentium Processor
: OPERATING SYSTEM: Microsoft Windows 95
: SOFTWARE: Microsoft Word 97
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/878,835A
: FILING DATE: June 19, 1997
: CLASSIFICATION: 800
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: AU PO 0605
: FILING DATE: 21 June 1996
: INFORMATION FOR SEQ ID NO: 7:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 10
: TYPE: Nucleic Acid
: STRANDEDNESS: Double
: TOPOLOGY: Linear
: US-08-878-835A-7

Query Match          100.0%; Score 6; DB 4; Length 10;
Best Local Similarity 100.0%; Pred. No. 1.6e+04;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ccgttc 6
   |||||
Db 6 CCGTTC 1

RESULT 6
US-07-974-447-10
```

```
: Sequence 10, Application US/07974447
: Patent No. 5436142
: GENERAL INFORMATION:
: APPLICANT: Wigler, Michael H
: APPLICANT: Lisitsyn, Nikolai
: TITLE OF INVENTION: A REPRESENTATIONAL APPROACH TO GENOMIC
: ANALYSIS
: NUMBER OF SEQUENCES: 16
: CORRESPONDENCE ADDRESSES:
: ADDRESSEE: FLEHR, HOBBACH, TEST, ALBRITTON & HERBERT
: STREET: 4 Embarcadero Center, Suite 3400
: CITY: San Francisco
: STATE: California
: COUNTRY: USA
: ZIP: 94111-4187
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patentln Release #1.0, Version #1.25
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/07/974,447
: FILING DATE: 12-NOV-1992
: CLASSIFICATION: 435
: ATTORNEY/AGENT INFORMATION:
: NAME: Rowland, Bertram I
: REGISTRATION NUMBER: 20,015
: REFERENCE/DOCKET NUMBER: A-57438/BIR CSHL-002
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (415) 781-1969
: TELEX: 910 277299
: INFORMATION FOR SEQ ID NO: 10:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 12 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: MOLECULE TYPE: cDNA
: US-07-974-447-10

Query Match          100.0%; Score 6; DB 1; Length 12;
Best Local Similarity 100.0%; Pred. No. 1.6e+04;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ccgttc 6
   |||||
Db 4 CCGTTC 9

RESULT 7
US-08-149-199-10
: Sequence 10, Application US/08149199
: Patent No. 5501964
: GENERAL INFORMATION:
: APPLICANT: Wigler, Michael H
: APPLICANT: Lisitsyn, Nikolai
: TITLE OF INVENTION: A REPRESENTATIONAL APPROACH TO GENOMIC
: ANALYSIS
: NUMBER OF SEQUENCES: 18
: CORRESPONDENCE ADDRESSES:
: ADDRESSEE: FLEHR, HOBBACH, TEST, ALBRITTON & HERBERT
: STREET: 4 Embarcadero Center, Suite 3400
: CITY: San Francisco
: STATE: California
: COUNTRY: USA
: ZIP: 941114187
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PCDOS/MSDOS
: SOFTWARE: Patentln Release #1.0, Version #1.25
```

```

; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/149,199
; FILING DATE: 9-NO. 5501964-93
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Rowland, Bertram I
; REGISTRATION NUMBER: 20,015
; REFERENCE/DOCKET NUMBER: A57438/BIR CSHL002
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 7811989
; TELEFAX: (415) 3983249
;
; TELEX: 910 277299
; INFORMATION FOR SEQ ID NO: 10:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 12 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: CDNA
; US-08-149-199-10

Query Match      100.0%; Score 6; DB 1; Length 12;
Best Local Similarity 100.0%; Pred. No. 1.6e+04;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 ccgttc 6
DB      4 cccgttc 9

RESULT      8
US-08-411-727-5/c
; Sequence 5, Application US/08411727
; Patent No. 5705161
; Patent No. 5705161 5683703
; GENERAL INFORMATION:
; APPLICANT: VAN DER LEY, Peter Andre
; APPLICANT: POOLMAN, Jan Theunis
; APPLICANT: HOOGERHOUT, Peter
; TITLE OF INVENTION: IMMUNOGENIC MENINGOCOCCAL LPS AND OTHER
; TITLE OF INVENTION: MEMBRANE VESICLES AND VACCINE THEREFROM
; NUMBER OF SEQUENCES: 30
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: YOUNG & THOMPSON
; STREET: 745 South 23rd Street, Suite 200
; CITY: Arlington
; STATE: Virginia
; COUNTRY: U.S.A.
; ZIP: 22202
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/411,727
; FILING DATE: 01-MAY-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: NL 9201716
; FILING DATE: 02-OCT-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: WO PCT/NL93/00163
; FILING DATE: 30-JUL-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: PATCH, Andrew J.
; REGISTRATION NUMBER: 32925
; REFERENCE/DOCKET NUMBER: BO 38275
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 703-521-2297
; TELEFAX: 703-685-0573
; TELEX: 248425 EMBON
```

```

; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 12 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; US-08-411-727-5

Query Match      100.0%; Score 6; DB 1; Length 12;
Best Local Similarity 100.0%; Pred. No. 1.6e+04;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 ccgttc 6
DB      11 cccgttc 6

RESULT      9
US-08-411-727-6
; Sequence 6, Application US/08411727
; Patent No. 5705161
; Patent No. 5705161 5683703
; GENERAL INFORMATION:
; APPLICANT: VAN DER LEY, Peter Andre
; APPLICANT: POOLMAN, Jan Theunis
; APPLICANT: HOOGERHOUT, Peter
; TITLE OF INVENTION: IMMUNOGENIC MENINGOCOCCAL LPS AND OTHER
; TITLE OF INVENTION: MEMBRANE VESICLES AND VACCINE THEREFROM
; NUMBER OF SEQUENCES: 30
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: YOUNG & THOMPSON
; STREET: 745 South 23rd Street, Suite 200
; CITY: Arlington
; STATE: Virginia
; COUNTRY: U.S.A.
; ZIP: 22202
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/411,727
; FILING DATE: 01-MAY-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: NL 9201716
; FILING DATE: 02-OCT-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: WO PCT/NL93/00163
; FILING DATE: 30-JUL-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: PATCH, Andrew J.
; REGISTRATION NUMBER: 32925
; REFERENCE/DOCKET NUMBER: BO 38275
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 703-521-2297
; TELEFAX: 703-685-0573
; TELEX: 248425 EMBON
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 12 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; US-08-411-727-6

Query Match      100.0%; Score 6; DB 1; Length 12;
Best Local Similarity 100.0%; Pred. No. 1.6e+04;
```

Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ccgttc 6  
|||||  
Db 2 CCgtTC 7

RESULT 10  
US-08-858-767-6  
; Sequence 6, Application US/08858767  
; Patent No. 5837468  
; GENERAL INFORMATION:  
; APPLICANT: WANG, Xun  
; APPLICANT: DUVICK, Jonathan P.  
; APPLICANT: BRIGGS, Steven P.  
; TITLE OF INVENTION: PCR-BASED CDNA SUBTRACTIVE CLONING  
; TITLE OF INVENTION: METHOD  
; NUMBER OF SEQUENCES: 39  
; CORRESPONDENCE ADDRESSES:  
; ADDRESSEE: Foley & Lardner  
; STREET: 3000 K Street, N.W., Suite 500  
; CITY: Washington  
; STATE: D.C.  
; COUNTRY: USA  
; ZIP: 20007-5109  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentin Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/858,767  
; FILING DATE: 19-MAY-1997  
; CLASSIFICATION: 435  
; PRIORITY APPLICATION DATA:  
; APPLICATION NUMBER: US 08/481,687  
; FILING DATE: 07-JUN-1995  
; ATTORNEY/AGENT INFORMATION:  
; NAME: BENT, Stephen A.  
; REGISTRATION NUMBER: 29,768  
; REFERENCE/DOCKET NUMBER: 33229/325/PIHI  
; TELEPHONE: (202)672-5300  
; TELEFAX: (202)672-5399  
; TELEX: 904136  
; INFORMATION FOR SEQ. ID NO: 6:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 12 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
US-08-858-767-6

Query Match 100.0%; Score 6; DB 2; Length 12;  
Best Local Similarity 100.0%; Pred. No. 1.6e+04;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ccgttc 6  
|||||  
Db 4 CCgtTC 9

RESULT 11  
US-08-858-767-8  
; Sequence 8, Application US/08858767  
; Patent No. 5837468  
; GENERAL INFORMATION:  
; APPLICANT: WANG, Xun  
; APPLICANT: DUVICK, Jonathan P.  
; APPLICANT: BRIGGS, Steven P.  
; TITLE OF INVENTION: PCR-BASED CDNA SUBTRACTIVE CLONING  
; TITLE OF INVENTION: METHOD

NUMBER OF SEQUENCES: 39  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Foley & Lardner  
STREET: 3000 K Street, N.W., Suite 500  
CITY: Washington  
STATE: D.C.  
COUNTRY: USA  
ZIP: 20007-5109  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/858,767  
FILING DATE: 19-MAY-1997  
CLASSIFICATION: 435  
PRIORITY APPLICATION DATA:  
APPLICATION NUMBER: US 08/481,687  
FILING DATE: 07-JUN-1995  
ATTORNEY/AGENT INFORMATION:  
NAME: BENT, Stephen A.  
REGISTRATION NUMBER: 29,768  
REFERENCE/DOCKET NUMBER: 33229/325/PIHI  
TELEPHONE: (202)672-5300  
TELEFAX: (202)672-5399  
TELEX: 904136  
INFORMATION FOR SEQ. ID NO: 8:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 12 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
US-08-858-767-8

Query Match 100.0%; Score 6; DB 2; Length 12;  
Best Local Similarity 100.0%; Pred. No. 1.6e+04;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ccgttc 6  
|||||  
Db 4 CCgtTC 9

RESULT 12  
US-08-863-028-6  
; Sequence 6, Application US/08863028  
; Patent No. 5853991  
; GENERAL INFORMATION:  
; APPLICANT: WANG, Xun  
; APPLICANT: DUVICK, Jonathan P.  
; APPLICANT: BRIGGS, Steven P.  
; TITLE OF INVENTION: PCR-BASED CDNA SUBTRACTIVE CLONING  
; TITLE OF INVENTION: METHOD  
; NUMBER OF SEQUENCES: 39  
; CORRESPONDENCE ADDRESSES:  
; ADDRESSEE: Foley & Lardner  
; STREET: 3000 K Street, N.W., Suite 500  
; CITY: Washington  
; STATE: D.C.  
; COUNTRY: USA  
; ZIP: 20007-5109  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentin Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/863,028  
; FILING DATE:  
; CLASSIFICATION: 435

PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US/08/858,767  
FILING DATE: 19-MAY-1997  
APPLICATION NUMBER: US 08/481,687  
FILING DATE: 07-JUN-1995  
ATTORNEY/AGENT INFORMATION:  
NAME: BENT, Stephen A.  
REGISTRATION NUMBER: 29,768  
REFERENCE/DOCKET NUMBER: 33229/325/PIHI  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (202)672-5300  
TELEFAX: (202)672-5399  
TELEX: 904136  
INFORMATION FOR SEQ ID NO: 6:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 12 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
US-08-863-028-6

Query Match 100.0%; Score 6; DB 2; Length 12;  
Best Local Similarity 100.0%; Pred. No. 1.6e+04;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ccgctc 6  
|||||  
DB 4 CCGTTC 9

RESULT 13  
US-08-863-028-8  
Sequence 8, Application US/08863028  
Patent No. 5853991  
GENERAL INFORMATION:  
APPLICANT: WANG, Xun  
APPLICANT: DIVICK, Jonathan P.  
APPLICANT: BRIGGS, Steven P.  
TITLE OF INVENTION: PCR-BASED CDNA SUBTRACTIVE CLONING  
NUMBER OF SEQUENCES: 39  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Foley & Lardner  
STREET: 3000 K Street, N.W., Suite 500  
CITY: Washington  
STATE: D.C.  
COUNTRY: USA  
ZIP: 20007-5109  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/863,028  
FILING DATE:  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US/08/858,767  
FILING DATE: 19-MAY-1997  
APPLICATION NUMBER: US 08/481,687  
FILING DATE: 07-JUN-1995  
ATTORNEY/AGENT INFORMATION:  
NAME: BENT, Stephen A.  
REGISTRATION NUMBER: 29,768  
REFERENCE/DOCKET NUMBER: 33229/325/PIHI  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (202)672-5300  
TELEFAX: (202)672-5399  
TELEX: 904136  
INFORMATION FOR SEQ ID NO: 8:  
SEQUENCE CHARACTERISTICS:

LENGTH: 12 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
US-08-863-028-8

Query Match 100.0%; Score 6; DB 2; Length 12;  
Best Local Similarity 100.0%; Pred. No. 1.6e+04;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ccgctc 6  
|||||  
DB 4 CCGTTC 9

RESULT 14  
US-09-115-061-10  
Sequence 10, Application US/09115061A  
Patent No. 6159713  
GENERAL INFORMATION:  
APPLICANT: Wigler, Michael  
APPLICANT: Lisitsyn, Nikolai  
TITLE OF INVENTION: A REPRESENTATIONAL APPROACH TO DNA ANALYSIS  
FILE REFERENCE: CSHL.002.030S  
CURRENT APPLICATION NUMBER: US/09/115,061A  
CURRENT FILING DATE: 1998-07-14  
EARLIER APPLICATION NUMBER: 08/478,242  
EARLIER FILING DATE: 1995-06-07  
EARLIER APPLICATION NUMBER: 08/149,199  
EARLIER FILING DATE: 1993-11-09  
EARLIER APPLICATION NUMBER: 07/974,447  
EARLIER FILING DATE: 1992-11-12  
NUMBER OF SEQ ID NOS: 20  
SOFTWARE: Patentin Ver. 2.0  
SEQ ID NO 10  
LENGTH: 12  
TYPE: DNA  
ORGANISM: Artificial Sequence  
FEATURE:  
OTHER INFORMATION: Description of Artificial Sequence:  
US-09-115-061-10

Query Match 100.0%; Score 6; DB 3; Length 12;  
Best Local Similarity 100.0%; Pred. No. 1.6e+04;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ccgctc 6  
|||||  
DB 4 ccgctc 9

RESULT 15  
US-09-261-079-10  
Sequence 10, Application US/09261079  
Patent No. 6277606  
GENERAL INFORMATION:  
APPLICANT: Wigler, Michael  
APPLICANT: Lisitsyn, Nikolai  
TITLE OF INVENTION: A REPRESENTATIONAL APPROACH TO DNA ANALYSIS  
FILE REFERENCE: CSHL.002.040S  
CURRENT APPLICATION NUMBER: US/09/261,079  
CURRENT FILING DATE: 1999-03-02  
EARLIER APPLICATION NUMBER: 08/478,242  
EARLIER FILING DATE: 1995-06-07  
EARLIER APPLICATION NUMBER: 07/974,447  
EARLIER FILING DATE: 1992-11-12  
NUMBER OF SEQ ID NOS: 20  
SOFTWARE: Patentin Ver. 2.0  
SEQ ID NO 10  
LENGTH: 12



; TYPE: DNA  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Description of Artificial Sequence:  
; OTHER INFORMATION: OLIGONUCLEOTIDE  
US-09-261-079-10

Query Match 100.0%; Score 6; DB 4; Length 12;  
Best Local Similarity 100.0%; Pred. No. 1.6e+04;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 ccgttc 6  
|||||  
Db 4 ccgttc 9

Search completed: July 29, 2002, 23:56:14  
Job time: 4755 sec



GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 29, 2002, 23:55:03 ; Search time 1921.77 Seconds  
(without alignments)  
65.335 Million cell updates/sec

Title: US-09-530-663b-16

Perfect score: 6

Sequence: 1 ccgttc 6

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 1797656 seqs, 10463268293 residues

Total number of hits satisfying chosen parameters: 3595312

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

GenEmbl: \*  
1: gb\_ba: \*  
2: gb\_htg: \*  
3: gb\_in: \*  
4: gb\_ov: \*  
5: gb\_ov: \*  
6: gb\_pat: \*  
7: gb\_ph: \*  
8: gb\_pl: \*  
9: gb\_pr: \*  
10: gb\_ro: \*  
11: gb\_sus: \*  
12: gb\_sy: \*  
13: gb\_un: \*  
14: gb\_vl: \*  
15: em\_ba: \*  
16: em\_fun: \*  
17: em\_hum: \*  
18: em\_in: \*  
19: em\_mu: \*  
20: em\_om: \*  
21: em\_or: \*  
22: em\_ov: \*  
23: em\_pat: \*  
24: em\_ph: \*  
25: em\_pl: \*  
26: em\_ro: \*  
27: em\_sts: \*  
28: em\_un: \*  
29: em\_vl: \*  
30: em\_htg\_hum: \*  
31: em\_htg\_inv: \*  
32: em\_htg\_other: \*  
33: em\_htgc\_inv: \*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result Query  
No. Score Match Length DB ID Description

C	1	6	100.0	10	6	AR070974	Sequence
C	2	6	100.0	10	6	AX043779	Sequence
C	3	6	100.0	10	6	AX152405	Sequence
C	4	6	100.0	10	6	AX152684	Sequence
C	5	6	100.0	10	6	AX301586	Sequence
C	6	6	100.0	12	6	A38146	Sequence 2
C	7	6	100.0	12	6	AR055101	Sequence
C	8	6	100.0	12	6	AR055103	Sequence
C	9	6	100.0	12	6	AR068442	Sequence
C	10	6	100.0	12	6	AR068444	Sequence
C	11	6	100.0	12	6	ARI21283	Sequence
C	12	6	100.0	12	6	I13336	Sequence
C	13	6	100.0	12	6	I19029	Sequence 10
C	14	6	100.0	14	6	BD000935	Method an
C	15	6	100.0	14	6	BD001364	Method an
C	16	6	100.0	14	6	I24766	Sequence 5
C	17	6	100.0	14	6	I24767	Sequence 6
C	18	6	100.0	14	6	I38426	Sequence 11
C	19	6	100.0	15	6	A01751	DNA fragment
C	20	6	100.0	15	6	A11090	Oligonucleo
C	21	6	100.0	15	6	A84556	Sequence 11
C	22	6	100.0	15	6	A89485	Sequence 16
C	23	6	100.0	15	6	AR033260	Sequence
C	24	6	100.0	15	6	AR033261	Sequence
C	25	6	100.0	15	6	AR113082	Sequence
C	26	6	100.0	15	6	AR113083	Sequence
C	27	6	100.0	15	6	ARI16412	Sequence
C	28	6	100.0	15	6	ARI21617	Sequence
C	29	6	100.0	15	6	ARI44754	Sequence
C	30	6	100.0	15	6	ARI66544	Sequence
C	31	6	100.0	15	6	ARI67451	Sequence
C	32	6	100.0	15	6	ARI76695	Sequence
C	33	6	100.0	15	6	AX108761	Sequence
C	34	6	100.0	15	6	AX196235	Sequence
C	35	6	100.0	15	6	I57489	Sequence 26
C	36	6	100.0	15	6	I57490	Sequence 27
C	37	6	100.0	16	6	AR008356	Sequence
C	38	6	100.0	16	6	AR030665	Sequence
C	39	6	100.0	16	6	AR053771	Sequence
C	40	6	100.0	16	6	ARI37835	Sequence
C	41	6	100.0	16	6	AX074441	Sequence
C	42	6	100.0	16	6	AX138221	Sequence
C	43	6	100.0	16	6	AX255602	Sequence
C	44	6	100.0	16	6	AX255636	Sequence
C	45	6	100.0	16	6	AX317642	Sequence

#### ALIGNMENTS

RESULT 1  
AR070974/c  
LOCUS AR070974  
DEFINITION Sequence 8 from patent US 5908978.  
ACCESSION AR070974  
VERSION AR070974.1 GI:7221862  
KEYWORDS  
SOURCE Unknown.  
ORGANISM Unknown.  
REFERENCE 1 (bases 1 to 10)  
AUTHORS Amerston,H.V., Wilcox,P., Sederoff,R.R., Kuhlman,E.George,  
O'Malley,D.M. and Grattapaglia,D.  
METHODS for within family selection of disease resistance in woody  
perennials using genetic markers  
Patent: US 5908978-A 8 01-JUN-1999;  
Location/Qualifiers  
1..10 /organism="unknown"

BASE COUNT 3 a 3 c 3 g 1 t  
ORIGIN

Query Match 100.0%; Score 6; DB 6; Length 10;  
 Best Local Similarity 100.0%; Pred. No. 1.7e+06;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 ccgttc 6  
 |||||  
 Db 6 CCGTTC 1

## RESULT 2

AX043779 10 bp DNA linear PAT 23-NOV-2000  
 LOCUS AX043779  
 DEFINITION Sequence 11 from Patent WO0065073.  
 ACCESION AX043779  
 VERSION AX043779.1 GI:11342383  
 KEYWORDS  
 SOURCE synthetic construct.  
 ORGANISM synthetic construct.  
 REFERENCE 1 (bases 1 to 10)  
 AUTHORS Martens,S. and Forkmann,G.  
 TITLE Genetic sequence which codes for the flavon synthase 11 enzyme and use of the same  
 JOURNAL Patent: WO 0065073-A 11 02-NOV-2000;  
 Martens, Stefan (DE) ; Forkmann, Gert (DE)  
 FEATURES  
 source 1..10  
 /organism="synthetic construct"  
 /db\_xref="taxon:32630"  
 /note="synthetisches Oligonukleotid"

BASE COUNT 0 a 4 c 4 g 2 t  
 ORIGIN

Query Match 100.0%; Score 6; DB 6; Length 10;  
 Best Local Similarity 100.0%; Pred. No. 1.7e+06;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 ccgttc 6  
 |||||  
 Db 3 CCGTTC 8

## RESULT 3

AX152405 10 bp DNA linear PAT 22-JUN-2001  
 LOCUS AX152405/c  
 DEFINITION Sequence 320 from Patent WO0138577.  
 ACCESION AX152405  
 VERSION AX152405.1 GI:14534056  
 KEYWORDS  
 SOURCE human.  
 ORGANISM Homo sapiens

REFERENCE 1 (bases 1 to 10)  
 AUTHORS Velculescu,V.E., Vogelstein,B. and Kinzler,K.W.  
 TITLE Human transcriptomes  
 JOURNAL Patent: WO 0138577-A 320 31-MAY-2001;  
 The Johns Hopkins University (US)  
 FEATURES  
 source 1..10  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"

BASE COUNT 3 a 2 c 5 g 0 t  
 ORIGIN

Query Match 100.0%; Score 6; DB 6; Length 10;  
 Best Local Similarity 100.0%; Pred. No. 1.7e+06;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 ccgttc 6  
 |||||

Db 9 CCGTTC 4

## RESULT 4

AX152684 10 bp DNA linear PAT 22-JUN-2001  
 LOCUS AX152684/c  
 DEFINITION Sequence 599 from Patent WO0138577.  
 ACCESION AX152684  
 VERSION AX152684.1 GI:14534335  
 KEYWORDS  
 SOURCE human.  
 ORGANISM Homo sapiens

REFERENCE 1 (bases 1 to 10)  
 AUTHORS Velculescu,V.E., Vogelstein,B. and Kinzler,K.W.  
 TITLE Human transcriptomes  
 JOURNAL Patent: WO 0138577-A 599 31-MAY-2001;  
 The Johns Hopkins University (US)  
 FEATURES  
 source 1..10  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"

BASE COUNT 3 a 1 c 6 g 0 t  
 ORIGIN

Query Match 100.0%; Score 6; DB 6; Length 10;  
 Best Local Similarity 100.0%; Pred. No. 1.7e+06;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 ccgttc 6  
 |||||  
 Db 8 CCGTTC 3

## RESULT 5

AX301586 10 bp DNA linear PAT 30-NOV-2001  
 LOCUS AX301586  
 DEFINITION Sequence 300 from Patent WO0185941.  
 ACCESION AX301586  
 VERSION AX301586.1 GI:17382669  
 KEYWORDS  
 SOURCE human.  
 ORGANISM Homo sapiens

REFERENCE 1 (sites)  
 AUTHORS Versteeg,R. and Caron,H.N.  
 TITLE Myc targets  
 JOURNAL Patent: WO 0185941-A 300 15-NOV-2001;  
 Academisch Ziekenhuis bij de Universiteit van Amsterdam (NL)  
 FEATURES  
 source 1..10  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"

BASE COUNT 1 a 3 c 2 g 4 t  
 ORIGIN

Query Match 100.0%; Score 6; DB 6; Length 10;  
 Best Local Similarity 100.0%; Pred. No. 1.7e+06;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 ccgttc 6  
 |||||  
 Db 2 CCGTTC 7

## RESULT 6

AX38146 12 bp DNA linear PAT 05-MAR-1997  
 LOCUS AX38146/c  
 DEFINITION Sequence 2 from Patent WO9408021.

ACCESSION A38146  
VERSION A38146.1 GI:2294752  
KEYWORDS  
SOURCE  
ORGANISM unidentified.  
REFERENCE 1 (bases 1 to 12)  
AUTHORS Van,D.L., Poolman,J.T. and Hoogerhout,P.  
TITLE IMMUNOGENIC MENINGOCOCCAL LPS AND OUTER MEMBRANE VESICLES AND VACCINE THEREFROM  
JOURNAL Patent: WO 9408021-A 2 14-APR-1994;  
NEDERLANDEN STAAT (NL)  
COMMENT Other publication AU 4835193 940426  
Other publication NO 951181 950601  
Other publication FI 951535 950601  
Other publication NL 9201716 940502  
Other publication JP 8501940T 960305.  
FEATURES  
source Location/Qualifiers  
1..12  
/organism="unidentified"  
/db\_xref="taxon:32644"  
/clone="PJBO07"  
BASE COUNT 3 a 3 c 5 g 1 t  
ORIGIN  
Query Match 100.0%; Score 6; DB 6; Length 12;  
Best Local Similarity 100.0%; Pred. No. 1.6e+06;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 ccgttc 6  
|||||  
Db 11 CCGTTC 6  
RESULT 7  
AR055101  
LOCUS AR055101 12 bp DNA linear PAT 29-SEP-1999  
DEFINITION Sequence 6 from patent US 5837468.  
ACCESSION AR055101  
VERSION AR055101.1 GI:5980678  
KEYWORDS  
SOURCE Unknown.  
ORGANISM Unknown.  
REFERENCE 1 (bases 1 to 12)  
AUTHORS Wang,X., Duvick,J.P. and Briggs,S.P.  
TITLE PCR-based cDNA subtractive cloning method  
JOURNAL Patent: US 5837468-A 6 17-NOV-1998;  
FEATURES  
source Location/Qualifiers  
1..12  
/organism="unknown"  
BASE COUNT 2 a 3 c 3 g 4 t  
ORIGIN  
Query Match 100.0%; Score 6; DB 6; Length 12;  
Best Local Similarity 100.0%; Pred. No. 1.6e+06;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 ccgttc 6  
|||||  
Db 4 CCGTTC 9  
RESULT 8  
AR055103  
LOCUS AR055103 12 bp DNA linear PAT 29-SEP-1999  
DEFINITION Sequence 8 from patent US 5837468.  
ACCESSION AR055103  
VERSION AR055103.1 GI:5980680  
KEYWORDS  
SOURCE Unknown.

ORGANISM Unknown.  
REFERENCE 1 (bases 1 to 12)  
AUTHORS Wang,X., Duvick,J.P. and Briggs,S.P.  
TITLE PCR-based cDNA subtractive cloning method  
JOURNAL Patent: US 5837468-A 8 17-NOV-1998;  
FEATURES  
source Location/Qualifiers  
1..12  
/organism="unknown"  
BASE COUNT 2 a 3 c 3 g 4 t  
ORIGIN  
Query Match 100.0%; Score 6; DB 6; Length 12;  
Best Local Similarity 100.0%; Pred. No. 1.6e+06;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 ccgttc 6  
|||||  
Db 4 CCGTTC 9  
RESULT 9  
AR068442  
LOCUS AR068442 12 bp DNA linear PAT 29-SEP-1999  
DEFINITION Sequence 6 from patent US 5853991.  
ACCESSION AR068442  
VERSION AR068442.1 GI:6000649  
KEYWORDS  
SOURCE Unknown.  
ORGANISM Unknown.  
REFERENCE 1 (bases 1 to 12)  
AUTHORS Wang,X., Duvick,J.P. and Briggs,S.P.  
TITLE PCR-based cDNA subtractive cloning method  
JOURNAL Patent: US 5853991-A 6 29-DEC-1998;  
FEATURES  
source Location/Qualifiers  
1..12  
/organism="unknown"  
BASE COUNT 2 a 3 c 3 g 4 t  
ORIGIN  
Query Match 100.0%; Score 6; DB 6; Length 12;  
Best Local Similarity 100.0%; Pred. No. 1.6e+06;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 ccgttc 6  
|||||  
Db 4 CCGTTC 9  
RESULT 10  
AR068444  
LOCUS AR068444 12 bp DNA linear PAT 29-SEP-1999  
DEFINITION Sequence 8 from patent US 5853991.  
ACCESSION AR068444  
VERSION AR068444.1 GI:6000651  
KEYWORDS  
SOURCE Unknown.  
ORGANISM Unknown.  
REFERENCE 1 (bases 1 to 12)  
AUTHORS Wang,X., Duvick,J.P. and Briggs,S.P.  
TITLE PCR-based cDNA subtractive cloning method  
JOURNAL Patent: US 5853991-A 8 29-DEC-1998;  
FEATURES  
source Location/Qualifiers  
1..12  
/organism="unknown"  
BASE COUNT 2 a 3 c 3 g 4 t  
ORIGIN

Query Match 100.0%; Score 6; DB 6; Length 12;  
 Best Local Similarity 100.0%; Pred. No. 1.6e+06;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ccgttc 6  
 |||||  
 Db 4 CCGTTC 9

## RESULT 11

LOCUS AR121283 12 bp DNA linear PAT 16-MAY-2001  
 DEFINITION Sequence 10 from patent US 6159713.  
 ACCESSION AR121283  
 VERSION AR121283.1 GI:14104859  
 KEYWORDS  
 SOURCE Unknown.  
 ORGANISM Unknown.  
 REFERENCE 1 (bases 1 to 12)  
 AUTHORS Wigler,M. and Lisitsyn,N.  
 TITLE Methods for producing probes capable of distinguishing DNA from related sources

JOURNAL Patent: US 6159713-A 10 12-DEC-2000;  
 FEATURES Location/Qualifiers  
 source 1..12  
 /organism="unknown"

BASE COUNT 2 a 3 c 3 g 4 t  
 ORIGIN

Query Match 100.0%; Score 6; DB 6; Length 12;  
 Best Local Similarity 100.0%; Pred. No. 1.6e+06;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ccgttc 6  
 |||||  
 Db 4 CCGTTC 9

## RESULT 12

LOCUS I13336 12 bp DNA linear PAT 26-JUL-1995  
 DEFINITION Sequence 10 from patent US 5436142.  
 ACCESSION I13336  
 VERSION I13336.1 GI:910677  
 KEYWORDS  
 SOURCE Unknown.  
 ORGANISM Unknown.  
 REFERENCE 1 (bases 1 to 12)  
 AUTHORS Wigler,M. and Lisitsyn,N.  
 TITLE Methods for producing probes capable of distinguishing variant genomic sequences

JOURNAL Patent: US 5436142-A 10 25-JUL-1995;  
 FEATURES Location/Qualifiers  
 source 1..12  
 /organism="unknown"

BASE COUNT 2 a 3 c 3 g 4 t  
 ORIGIN

Query Match 100.0%; Score 6; DB 6; Length 12;  
 Best Local Similarity 100.0%; Pred. No. 1.6e+06;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ccgttc 6  
 |||||  
 Db 4 CCGTTC 9

## RESULT 13

FT I19029

LOCUS I19029 12 bp DNA linear PAT 07-OCT-1996  
 DEFINITION Sequence 10 from patent US 5501964.  
 ACCESSION I19029  
 VERSION I19029.1 GI:1599384  
 KEYWORDS  
 SOURCE Unknown.  
 ORGANISM Unknown.  
 REFERENCE 1 (bases 1 to 12)  
 AUTHORS Wigler,M. and Lisitsyn,N.  
 TITLE Methods for producing probes capable of distinguishing DNA from related sources

JOURNAL Patent: US 5501964-A 10 26-MAR-1996;  
 FEATURES Location/Qualifiers  
 source 1..12  
 /organism="unknown"

BASE COUNT 2 a 3 c 3 g 4 t  
 ORIGIN

Query Match 100.0%; Score 6; DB 6; Length 12;  
 Best Local Similarity 100.0%; Pred. No. 1.6e+06;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ccgttc 6  
 |||||  
 Db 4 CCGTTC 9

## RESULT 14

LOCUS BD000935 14 bp RNA linear PAT 31-JAN-2002  
 DEFINITION Method and reagent for inhibiting viral replication.  
 ACCESSION BD000935  
 VERSION BD000935.1 GI:18625494  
 KEYWORDS JP 2000342285-A/95.  
 SOURCE synthetic construct.  
 ORGANISM artificial sequence.  
 REFERENCE 1 (bases 1 to 14)  
 AUTHORS Draper,K.G., Dadykiz,L.W., Macswigen,J.A., Maysejck,D.G., Holesek,J.J., and Mamone,A.J.  
 TITLE Method and reagent for inhibiting viral replication  
 JOURNAL Patent: JP 2000342285-A 95 12-DEC-2000;  
 COMMENT RIBOZYME PHARMACEUTICALS INC

OS Artificial Sequence  
 PN JP 2000342285-A/95  
 PD 12-DEC-2000  
 PF 01-MAY-2000 JP 2000132616

PR 11-MAY-1992 US 07/882689,14-MAY-1992 US 07/882712 PR  
 14-MAY-1992 US 07/882713,14-MAY-1992 US 07/882714 PR  
 14-MAY-1992 US 07/882823,14-MAY-1992 US 07/882824 PR  
 14-MAY-1992 US 07/882868,14-MAY-1992 US 07/882868 PR  
 14-MAY-1992 US 07/882889,14-MAY-1992 US 07/882921 PR  
 14-MAY-1992 US 07/882922,14-MAY-1992 US 07/883823 PR  
 14-MAY-1992 US 07/883849,14-MAY-1992 US 07/884073 PR  
 14-MAY-1992 US 07/884074,14-MAY-1992 US 07/884333 PR  
 14-MAY-1992 US 07/884422,14-MAY-1992 US 07/884431 PR  
 14-MAY-1992 US 07/884436,14-MAY-1992 US 07/884521 PR  
 31-JUL-1992 US 07/923738,26-AUG-1992 US 07/935854 PR  
 26-AUG-1992 US 07/936086,18-SEP-1992 US 07/948359 PR  
 15-OCT-1992 US 07/963322,07-DEC-1992 US 07/987129 PR  
 07-DEC-1992 US 07/987130,07-DEC-1992 US 07/987133 PR  
 KENNETH G DRAPER, LEC W DADYKIZ, JAMES A MACSWIGEN, PI DENNIS G  
 MAYSEJCK,  
 PI JAMES J HOLESEK, ANTHONY J MAMONE  
 PC C12N15/09, C12N5/10, C12N7/00, C12N9/22//C12N5/10, C12R1.91), PC  
 C12N15/00,  
 CC C12N5/00, (C12N5/00, C12R1.91)

PC C12N5/00, (C12N5/00, C12R1.91)  
 FH Key location/Qualifiers  
 FT source 1..14  
 /organism="Artificial Sequence".

FEATURES  
source Location/Qualifiers  
1..14  
/organism="synthetic construct"  
/db\_xref="taxon:32630"

BASE COUNT 1 a 7 c 3 g 3 t

ORIGIN

Query Match 100.0%; Score 6; DB 6; Length 14;  
Best Local Similarity 100.0%; Pred. No. 1.6e+06;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 ccgttc 6  
|||||  
Db 4 CCGTTC 9

RESULT 15  
BD001364 14 bp RNA linear PAT 31-JAN-2002  
LOCUS BD001364  
DEFINITION Method and reagent for inhibiting viral replication.  
ACCESSION BD001364  
VERSION BD001364.1 GI:18625923  
KEYWORDS JP 2000342286-A/95.  
SOURCE synthetic construct.  
ORGANISM synthetic construct.  
REFERENCE 1 (bases 1 to 14)  
AUTHORS Draper,K.G., Dadykiz,L.W., Macswigen,J.A., Maysejak,D.G.,  
Holesek,J.J. and Mamone,A.J.  
TITLE Method and reagent for inhibiting viral replication  
JOURNAL Patent: JP 2000342286-A 95 12-DEC-2000;  
RIBOZYME PHARMACEUTICALS INC  
COMMENT OS Artificial Sequence  
PN JP 2000342286-A/95  
PD 12-DEC-2000  
PF 01-MAY-2000 JP 2000132651  
PR 11-MAY-1992 US 07/882689,14-MAY-1992 US 07/882712 PR  
14-MAY-1992 US 07/882713,14-MAY-1992 US 07/882714 PR  
14-MAY-1992 US 07/882823,14-MAY-1992 US 07/882824 PR  
14-MAY-1992 US 07/882866,14-MAY-1992 US 07/882868 PR  
14-MAY-1992 US 07/882889,14-MAY-1992 US 07/882921 PR  
14-MAY-1992 US 07/882922,14-MAY-1992 US 07/883823 PR  
14-MAY-1992 US 07/883849,14-MAY-1992 US 07/884073 PR  
14-MAY-1992 US 07/884074,14-MAY-1992 US 07/884333 PR  
14-MAY-1992 US 07/884422,14-MAY-1992 US 07/884431 PR  
14-MAY-1992 US 07/884436,14-MAY-1992 US 07/884521 PR  
31-JUL-1992 US 07/923738,26-AUG-1992 US 07/935854 PR  
26-AUG-1992 US 07/936086,18-SEP-1992 US 07/948359 PR  
15-OCT-1992 US 07/963322,07-DEC-1992 US 07/987129 PR  
07-DEC-1992 US 07/987130,07-DEC-1992 US 07/987133 PI  
KENNETH G DRAPER,LEC W DADYKIZ,JAMES A MACSWIGEN, PI DENNIS G  
MAYSEJAK,

PI JAMES J HOLESSEK,ANTHONY J MAMONE  
PC C12N15/09,C12N5/10,C12N7/00//A61K38/43,A61K39/125,A61K39/13,  
PC A61K39/135,  
PC A61K39/145,A61K39/21,A61K39/23,A61K39/245,A61K39/29,A61K48/00,  
PC A61P1/16,  
PC A61P31/14,A61P31/16,A61P31/18,A61P31/22,A61P35/02,C12O1/68,PC  
(C12N15/09,C12R1/93),C12N15/00,C12N5/00,A61K37/48,(C12N15/00,PC  
C12R1/93)

FEATURES  
source Location/Qualifiers  
1..14  
/organism="synthetic construct"  
/db\_xref="taxon:32630"

BASE COUNT 1 a 7 c 3 g 3 t

ORIGIN

Query Match 100.0%; Score 6; DB 6; Length 14;  
Best Local Similarity 100.0%; Pred. No. 1.6e+06;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 ccgttc 6  
|||||  
Db 4 CCGTTC 9

Search completed: July 29, 2002, 23:55:06  
Job time: 7022 sec

